

# Fusion of Hyperspectral and LIDAR Remote Sensing Data for Classification of Complex Forest Areas

Michele Dalponte, *Student Member, IEEE*, Lorenzo Bruzzone, *Senior Member, IEEE*, and Damiano Gianelle

**Abstract**—In this paper, we propose an analysis on the joint effect of hyperspectral and light detection and ranging (LIDAR) data for the classification of complex forest areas. In greater detail, we present: 1) an advanced system for the joint use of hyperspectral and LIDAR data in complex classification problems; 2) an investigation on the effectiveness of the very promising support vector machines (SVMs) and Gaussian maximum likelihood with leave-one-out-covariance algorithm classifiers for the analysis of complex forest scenarios characterized from a high number of species in a multisource framework; and 3) an analysis on the effectiveness of different LIDAR returns and channels (elevation and intensity) for increasing the classification accuracy obtained with hyperspectral images, particularly in relation to the discrimination of very similar classes. Several experiments carried out on a complex forest area in Italy provide interesting conclusions on the effectiveness and potentialities of the joint use of hyperspectral and LIDAR data and on the accuracy of the different classification techniques analyzed in the proposed system. In particular, the elevation channel of the first LIDAR return was very effective for the separation of species with similar spectral signatures but different mean heights, and the SVM classifier proved to be very robust and accurate in the exploitation of the considered multisource data.

**Index Terms**—Data fusion, forestry, hyperspectral images, light detection and ranging (LIDAR) data, multisensor classification.

## I. INTRODUCTION

**F**OREST preservation and management are important and complex processes, which have significant implications on the environment (e.g., protection of biological diversity and climate mitigation) and on the economy (e.g., estimation of timber volume for commercial usage). An efficient prevention and management policy requires a detailed knowledge of species composition, distribution, and density. However, the assessment of the distribution of tree species in large forests by a ground inventory is a difficult and time-consuming task. Remote sensing is a very useful technology to perform such kind of study. This technology, if properly integrated with automatic processing techniques, allows the analysis of large areas in a fast and accurate way. Several studies have been carried out in this field, analyzing the potentialities of different remote sensing sensors, including passive multispectral and hyperspectral sensors, as

well as active light detection and ranging (LIDAR) and synthetic aperture radar (SAR) systems (e.g., [1]–[36]). All these sensors, with their different peculiarities and characteristics, can provide different information about the analyzed forest, allowing to reach different targets such as classification of tree species or estimation of biophysical parameters.

Standard passive multispectral sensors (like the Thematic Mapper of the Landsat satellites) have been widely used in the past years for forest classification and analysis. In the literature, several studies are present on both classification and estimation of forest parameters (e.g., [1]–[4]). Regarding classification, due to the different spectral and geometrical characteristics of multispectral sensors available, it is possible to find works that analyze the problem with different levels of geometrical detail. Regarding low-resolution multispectral data, the analysis is generally limited to the discrimination between forested and nonforested areas (see, for example, [5]). With medium-resolution sensors, the level of detail can be increased, and thus, the analysis can be focused on more specific classes, like in the study presented in [6] where, using Landsat Enhanced Thematic Mapper Plus (ETM+) images, eight different vegetation classes are analyzed. High geometrical resolution multispectral sensors (e.g., Quickbird, Ikonos, and SPOT5) allow a more detailed geometrical analysis considering the high spatial resolution, but due to the poor spectral information acquired by these sensors, they do not allow a detailed analysis of tree species. As an example, Kosaka *et al.* [7] analyze six forest types using Quickbird images and Wang *et al.* [8] distinguish three kinds of mangrove using Ikonos and Quickbird data.

However, although significant results in forest analysis can be obtained with these kinds of data, in forest characterized by a high number of similar tree species, these sensors do not allow a detailed analysis of the different forest species, as they acquire information in a relatively small number of bands with large spectral intervals. The new generation of passive hyperspectral sensors, due to their ability to make a dense sampling of the spectral signature, can instead collect a valuable information for a detailed classification and analysis of similar forest types. In particular, these data can be used in a wide range of different analyses of forest environments. Several studies have addressed the capability of hyperspectral data to estimate particular biophysical parameters like chlorophyll concentration or biomass volume (e.g., [9]–[11]). Concerning classification problems, hyperspectral images have been used in a wide number of forest applications, ranging from general cases focusing on the discrimination between forest and other land covers, to a more detailed analysis dealing with the distinction of different tree species (e.g., [6], [12]–[14]). In [6], for example, Goodenough *et al.* present an interesting analysis comparing

Manuscript received July 6, 2007; revised November 22, 2007.

M. Dalponte is with the Department of Information and Communication Technology, University of Trento, 38050 Trento, Italy, and also with the Centro di Ecologia Alpina, 38100 Trento, Italy (e-mail: michele.dalponte@dit.unitn.it).

L. Bruzzone is with the Department of Information and Communication Technology, University of Trento, 38050 Trento, Italy (e-mail: lorenzo.bruzzone@dit.unitn.it).

D. Gianelle is with the Centro di Ecologia Alpina, 38100 Trento, Italy (e-mail: gianelle@cealp.it).

Digital Object Identifier 10.1109/TGRS.2008.916480

classification results on a forest area obtained with three different sensors, two multispectral (i.e., the Landsat-7 ETM+ and the EO-1 ALI), and one hyperspectral (i.e., the EO-1 Hyperion). The results of this study confirmed that, with hyperspectral data, it is possible to reach much higher classification accuracies than with multispectral images. In [13], Clark *et al.* studied seven deciduous tree species with the HYDICE sensor, using three different classifiers, reaching accuracies on the order of 90%. In [14], Leckie *et al.* used CASI hyperspectral images to separate five different coniferous species, demonstrating the high importance of these kinds of data in classification of similar tree species. In [12], Martin *et al.* separated 11 forest classes using Airborne Visible/Infrared Imaging Spectrometer data.

Active SAR and LIDAR remote sensing sensors are also widely used in forest analysis. SAR system is an important source of information for studies on forest environments. With SAR data, it is possible to estimate a wide range of forest parameters, ranging from structure to biophysical indexes, like forest fuel load (e.g., [15]–[17]). In the classification domain, SAR data are mainly used for the separation of forested from nonforested areas [18] or in problems where classification is connected with tree parameters. In this context, Lee *et al.* [19] classify different stages of the age of coniferous and deciduous trees using L-band polarimetric interferometric SAR data. Ranson *et al.* [20] present a similar work, studying Siberian trees, dividing the vegetation in four classes: young deciduous, old deciduous, young conifer, and old conifer. In [21], Saatchi and Rignot classify seven different vegetation classes (out of a total of eight) using Jet Propulsion Laboratory aircraft SAR data.

The use of LIDAR sensors is increasing in the context of forest applications. LIDAR is an effective information source for studies related to tree height, forest structure, biomass, and all the parameters that are mainly related to the vertical dimension of the scene under analysis (e.g., [22]–[25]). LIDAR potentially allows a very precise and detailed analysis of different forest parameters. For example, in [25], Andersen *et al.* study the potentialities of LIDAR in the estimation of some forest canopy fuel parameters, finding high correlation between LIDAR data and biophysical parameters. Some studies have also been done in using LIDAR data in classification problems, in particular in cases where a reduced number of classes are investigated, such as the case of discrimination between deciduous and conifer trees (e.g., [26]–[28]). In [26], Brennan and Webster present a study with nine classes, obtaining high classification accuracies for all classes and emphasizing that LIDAR data can be very effective in the distinction between coniferous and deciduous trees. In [28], Holmgren and Persson identify species of individual trees using high-density airborne laser scanner data characterizing the structure and the shape of different tree species.

The high number of remote sensing sensors available in these last years, as well as the possibility to have images acquired by different sensors on the same area, has resulted in several studies on the use of multisensor information for forest applications. In this context, many papers have been published on the joint use of multispectral (or hyperspectral) images and SAR data (e.g., [29] and [30]). Recently, some works have also addressed the joint use of LIDAR and other active

and passive sensors in forest parameter estimation problems (e.g., [31]–[34]). For example, in [34], Hyde *et al.* describe the results of an analysis on forest structure using four different sensors (i.e., LIDAR, SAR, Landsat ETM+, and Quickbird), underlining that, for the estimation of forest parameters, the combination of LIDAR and ETM+ data achieves good accuracy. Concerning classification problems, Simental *et al.* [35] explore the joint use of hyperspectral and LIDAR data for the separation of vegetation classes, underlining that LIDAR can be very useful in the separation of shrubs from trees. In [36], Lemp and Weidner exploit hyperspectral and LIDAR data for the classification of urban areas, using LIDAR for the segmentation of the scene, and then hyperspectral data for the classification of the resulting regions. In [37], Mundt *et al.* present a study on the joint use of hyperspectral and LIDAR data for the classification of sagebrush distribution, reporting accuracies of about 80%. In [38], Sugumaran and Voss address the joint use of hyperspectral and LIDAR data for the identification of tree species in an urban environment, showing the effectiveness of LIDAR bands in the classification phase. Other studies exploit LIDAR data in the preprocessing phase. For example, in [39], Perry *et al.* use the digital terrain model (DTM) derived from LIDAR in the phase of geometric correction of hyperspectral images.

All the aforementioned papers indicate a good complementary relationship between hyperspectral and LIDAR data, as they contain very different information: hyperspectral images provide a detailed description of the spectral signatures of classes but no information on the height of ground covers, whereas LIDAR data give detailed information about the height but no information on the spectral signatures. However, most of the studies do not approach the integration of LIDAR and hyperspectral signals from a real data fusion perspective but address the problem in terms of separate use of these information sources in different processing phases. In this scenario, at the present, only very few investigations have been carried out on both the design of advanced classification systems capable of properly exploiting the complementary information present in these data and the possibility to jointly use LIDAR and hyperspectral data for classification of complex forest areas in presence of many tree species.

In this paper, we address the above issues by proposing an advanced classification system for the joint exploitation of LIDAR and hyperspectral data and by studying the importance of LIDAR data when fused with hyperspectral images in solving complex forest classification problems. The main motivation of this paper is that, at the present time, it is becoming more common to acquire both LIDAR and hyperspectral data on forest areas. Generally, these data are used separately; in particular, hyperspectral data are exploited for forest classification and LIDAR data for forest parameter estimation. However, the availability of both data can be properly exploited in a data fusion framework both at the classification and the estimation level. In this paper, we focus our attention on the classification problem. The main contributions of this paper to the literature are as follows.

- 1) Definition of an advanced system for the joint use of hyperspectral and LIDAR data in classification of complex forest areas. In particular, the proposed system can properly manage: a) the hyperdimensionality of the

TABLE I  
DISTRIBUTION OF GROUND TRUTH SAMPLES AMONG INVESTIGATED CLASSES

CLASS NAME	GROUND TRUTH SAMPLES	CLASS NAME	GROUND TRUTH SAMPLES	CLASS NAME	GROUND TRUTH SAMPLES
<i>Acer campestre</i>	170	<i>Juglans regia</i>	1573	<i>Quercus rubra</i>	1137
<i>Acer negundo</i>	48	<i>Morus sp.</i>	164	<i>Robinia pseudoacacia</i>	1008
<i>Alnus glutinosa</i>	507	<i>Platanus hybrida</i>	2048	<i>Rubus</i>	661
<i>Carpinus betulus</i>	910	<i>Populus canescens</i>	244	Shadows	290
<i>Corylus avellana</i>	58	<i>Populus hybrida</i>	211	Snags	205
<i>Fraxinus angustifolia</i>	787	<i>Prunus avium</i>	261	<i>Tilia cordata</i>	507
Grassland	496	<i>Quercus cerris</i>	1796	<i>Ulmus minor</i>	403
<i>Juglans nigra</i>	1283	<i>Quercus robur</i>	2049		

feature vector intrinsic in hyperspectral data; b) the different statistical properties of hyperspectral and LIDAR data; and c) the complementary role that LIDAR data can play with respect to hyperspectral data for the discrimination of some important forest species.

- 2) Investigation on the effectiveness of the very promising distribution-free support vector machines (SVMs) and the parametric Gaussian maximum likelihood with leave-one-out-covariance algorithm (GML-LOOC) classifiers in the analysis of complex forest scenarios characterized by a high number of species in a multisource framework.
- 3) Analysis on the effectiveness of different LIDAR returns and channels (elevation and intensity) for increasing the classification accuracy obtained with hyperspectral images, particularly in relation to the discrimination of very similar classes.

The proposed system was tested on a dense forest area characterized by a very high number of complex tree species (i.e., 19 species). In the experiments, we considered airborne hyperspectral images and LIDAR data with a very high geometrical resolution (1 m) and a density higher than five points per square meter, respectively. The results obtained confirm the effectiveness of the proposed system and achieve interesting conclusions on the importance of the joint use of LIDAR and hyperspectral data in forest classification.

This paper is organized into six sections. Section II describes the data set used in our analysis, whereas Section III presents the problem definition and the architecture of the proposed system, as well as the main preprocessing techniques adopted. The classification methods investigated in the proposed system are analyzed in Section IV. Section V describes and discusses the experimental results obtained. Finally, Section VI draws the conclusion of this paper.

## II. DATA SET DESCRIPTION

The study area selected is a complex forest scene that corresponds to the natural reserve of the “*Bosco della Fontana*” in the Po Plain near the city of Mantua, Italy. The central point of the area has the following coordinates: 45° 12' 1.68" N, 10° 44' 35.53" E. The topography of this area is almost perfectly flat, and it extends across an area of approximately 230 ha. This area represents one of the best preserved forest relicts on the Po Plain. Due to the absence of a significant human impact in the last century, this area has the following interesting properties: 1) it contains a high number of vegetation species (more than 20); 2) it consists of several similar tree species, including

*Quercus cerris*, *Quercus robur*, and *Quercus rubra*; and 3) it does not exhibit a preordered spatial tree distribution.

We investigated 19 different tree species, to which we added a further four classes to have an exhaustive representation of land covers of the whole area analyzed. In total, 23 classes were represented (see Table I for a complete description of the classes investigated). It is worth noting that among the 19 tree species under analysis, there are classes belonging to the same family, which have very similar spectral signatures. Another important consideration with respect to this data set is that, from the analyzed area, not all the vegetation classes have the same relative frequency, and that there are some dominant species (e.g., *Carpinus betulus*, *Quercus cerris*, *Quercus robur*, and *Quercus rubra*).

The hyperspectral and LIDAR data (see Fig. 1) were acquired simultaneously on June 28, 2006 between 9:04 A.M. and 9:36 A.M. The hyperspectral data consist of six partially overlapping images acquired by an AISA Eagle sensor in 126 spectral bands, ranging from 400 to 990 nm, with a spectral resolution of about 4.5 nm and a spatial resolution of 1 m. The flight direction of the plane was the same for all the six images (from East to West), and the flight height was consistent, at approximately 750 m. The LIDAR data were acquired by a sensor Optech ALTM 3100, with a mean density of 5.6 points per square meter. The laser pulse wavelength and the laser repetition rate were 1064 nm and 100 kHz, respectively. The data used in our investigation refer to the first four LIDAR returns, in particular, the elevation and the intensity channels of each return. The total number of LIDAR points per return is as follows: 20 271 067 points for the first return, 5 096 256 for the second, 1 110 799 for the third, and 85 741 for the fourth. A DTM of the investigated area with a spatial resolution of 1 m was extracted from the LIDAR data.

The ground truth samples (approximately 550 points) were collected with a ground survey in autumn 2006. Samples were collated on a laptop within an orthophoto (with a geometrical resolution of 20 cm) of the area analyzed according to ground observations. We extracted these points from the entire study area, thus, ensuring a precise matching between the ground observations and the aerial ones (e.g., we considered trees near roads and grassland). The samples were collected on the basis: 1) of the species (the ground truth is exhaustive, i.e., it represents all the species present in the area; furthermore, it takes into account the relative frequency of each class) and 2) of the spatial distribution (samples have a uniform distribution all over the scene). All points were then converted to region of interests on the coregistered hyperspectral and LIDAR data and used for the generation of the

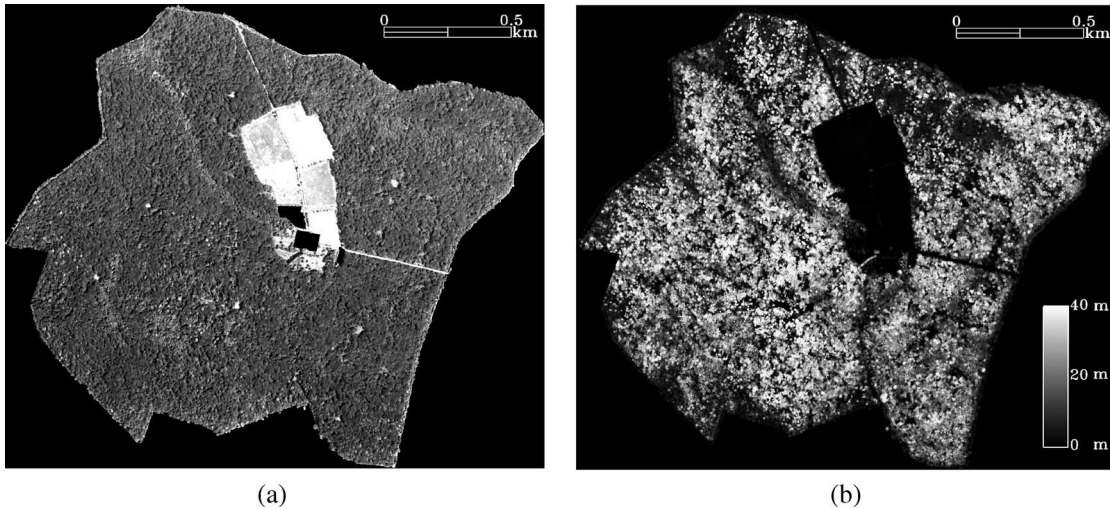


Fig. 1. Example of images used in the experiments. (a) Channel 34 (550 nm) of the hyperspectral image. (b) DCM of the analyzed area.

training and test sets. The total number of ground truth samples (16816 pixels) represents about 0.7% of the whole investigated area.

### III. PROBLEM DEFINITION AND SYSTEM ARCHITECTURE

#### A. Problem Definition

Generally, the analysis of large forest areas with hyperspectral scanners (usually characterized from a relatively small field of view) requires the acquisition of different images that are then integrated according to a mosaic procedure. In this context, let us consider a series of  $M$  hyperspectral images  $\mathbf{X}_i$  ( $i = 1, \dots, M$ ) acquired in partially overlapping portions of the investigated area, and a LIDAR image  $\mathbf{L}$  taken simultaneously with the hyperspectral ones. Let  $\mathbf{H}$  be the radiometric normalized mosaic of these images, and  $\mathbf{H}_{nr}$  be the corresponding noise-reduced hyperspectral image.  $\mathbf{X}_i$  ( $i = 1, \dots, M$ ),  $\mathbf{H}$ , and  $\mathbf{H}_{nr}$  are  $n$ -dimensional images, where  $n$  is the total number of spectral bands. Let  $\mathbf{L}$  denote the LIDAR interpolated image that consists of the elevation and the intensity channels of the first  $m$  LIDAR returns. The total number of bands of  $\mathbf{L}$  is  $2 * m$ , due to the fact that for each return, we have both elevation  $\mathbf{E}$  and intensity  $\mathbf{I}$  image (i.e.,  $\mathbf{L} = \mathbf{E} \cup \mathbf{I}$ ). Thus,  $\mathbf{E}$  and  $\mathbf{I}$  are  $m$ -dimensional images, representing the elevation and the intensity of the first  $m$  LIDAR returns, respectively. Let  $\mathbf{x}_p$  be the  $q$ -dimensional feature vector that represents the  $p$ th pattern in input to the classifier. Finally, let  $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$  be the set of the  $K$  land-cover classes in the considered classification problem, with  $\omega_i$  the  $i$ th class.

As stated in Section I, we focus on a specific problem: the fusion of hyperspectral and LIDAR data for classification of trees species. To reach this objective, we propose a system based on an architecture that processes both hyperspectral and LIDAR data, exploiting the complementary role that these data can play. The architecture of the proposed system (with the aforementioned notation) is shown in Fig. 2. In the following, we present, in detail, the different parts of the system.

#### B. System Architecture

The analysis of two different kinds of data (hyperspectral and LIDAR) requires the use of two different preprocessing

schemes. For hyperspectral data, according to what was previously described, it is necessary to mosaic various images to achieve coverage of the whole site. Before this phase, a relative radiometric normalization should be applied to the single images to obtain a uniform mosaic image. Several normalization algorithms have been proposed in literature (e.g., [40] and [41]). Since the investigated area is almost perfectly flat, and the data were acquired in a reduced interval of time (about 30 min), it is reasonable to assume that all six hyperspectral images were taken under the same illumination conditions. Therefore, and taking into account that in the classification phase we use a supervised classification system, we applied a relative radiometric normalization to the images without any specific atmospheric correction. In greater detail, we adopted a simple linear normalization based on the mean-standard deviation algorithm [40].

After creating the mosaic, we coregistered the hyperspectral data to the LIDAR images, using approximately 75 ground control points distributed across the entire image. In particular, to warp the image, we selected a polynomial transformation of third order and a nearest-neighbor resampling of the pixels. The root-mean-square error resulting after the coregistration phase was 0.76. The hyperspectral data were then denoised with a simple low-pass filter with a window size of  $3 \times 3$  pixels. In the previously published literature, several studies pointed out the usefulness of this operation (e.g., [42] and [43]). In our case, given the high geometrical resolution of the images, the spatial degradation involved by the filter is acceptable with respect to both the reduction of the noise present in the images and the expected increase in class separability [42].

From a methodological viewpoint, the automatic analysis of hyperspectral data in the presence of a high number of forest classes is not a trivial task. In particular, the complexity can be attributed to: 1) the high computational cost; 2) the need of advanced classification systems capable of adequately modeling the nonlinear hyperdimensional discrimination functions associated with the presence of many tree species; and 3) the curse of dimensionality. In the context of a supervised classification, one of the main difficulties is related to the usually small ratio between the number of available training samples and the number of features (Hughes phenomenon [44]) that makes it difficult (or impossible) to estimate the parameters

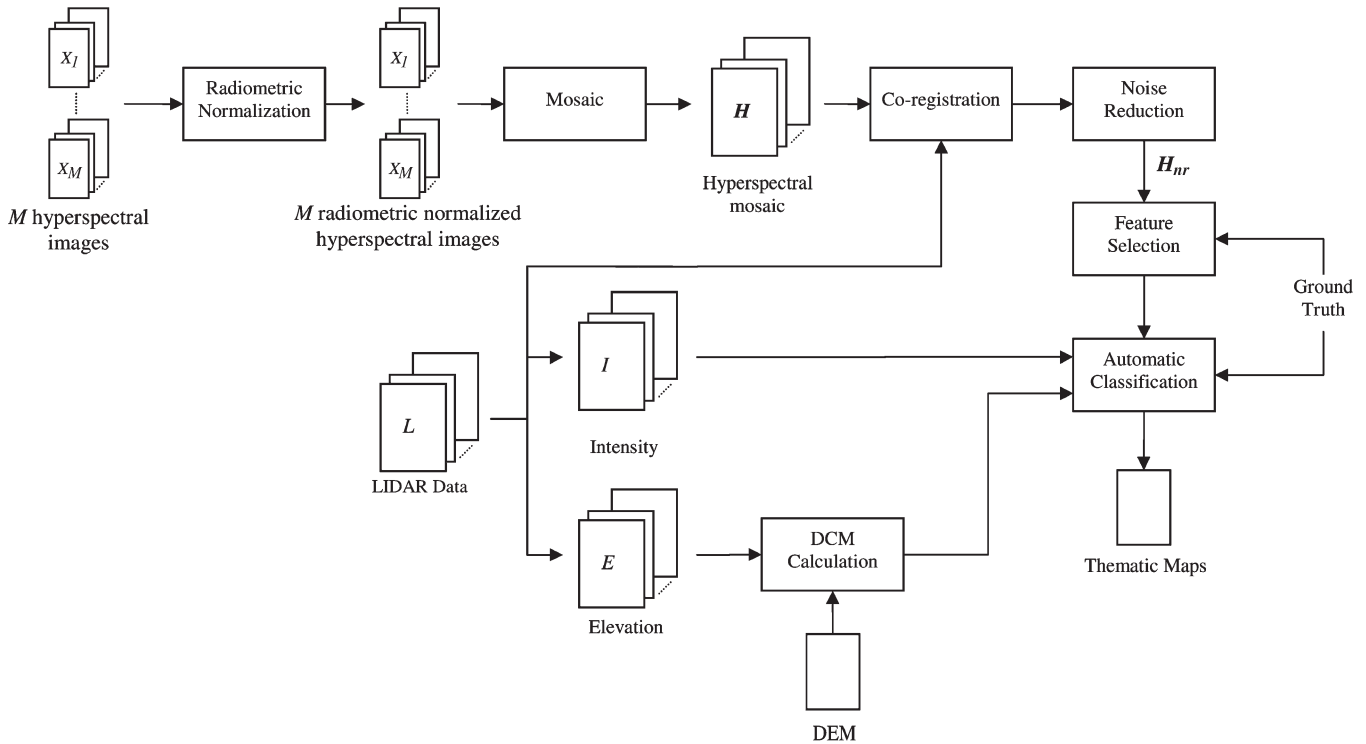


Fig. 2. Architecture of the system developed.

of the classifier [e.g., with the Gaussian maximum-likelihood (GML) algorithm].

A possible solution to this problem is to use a feature selection technique. The rationale of this approach is to reduce the number of features by selecting a representative subset of the original spectral channels. A feature selection technique is made up of a search strategy and of a separability criterion. In the literature, several algorithms have been proposed for both these tasks. Concerning the search strategy, we can find optimal procedures (e.g., branch and bound [43]), which allow us to identify the subset of features that maximizes the separability criterion, or suboptimal ones (e.g., sequential forward floating selection (SFFS) [45], steepest ascent [46]), which find a suboptimal solution with a reduced computational cost. For our study, we adopted the SFFS algorithm, which is widely used in the literature as it provides solutions reasonably close to the optimal one. Regarding the separability criterion, several measures have been presented in the literature, including the Bhattacharyya distance, the Jeffreys–Matusita distance, and transformed divergence [43], [47]. We selected the Jeffreys–Matusita distance, which is associated to the Chernoff upper bound to the Bayesian error also in the multiclass case [48].

For the LIDAR analysis, we rasterized the raw data (corresponding to the LAS format) of all the returns. The elevation and the intensity channels were converted into a raster image with a spatial resolution of 1 m, assigning to each pixel the mean value of points within the corresponding area on the ground. The few pixels with missing data in the first return were replaced by a linear interpolation, whereas no interpolation was applied to the other returns. After this phase, to determine the height of vegetation with respect to the ground, we extracted the digital canopy model (DCM) by subtracting the DTM to

the elevation channel of the LIDAR return. This procedure was applied to the elevation band of all four LIDAR returns.

#### IV. CLASSIFICATION TECHNIQUES

In the definition of the proposed system, we analyzed and compared two advanced classification techniques, specifically suitable to the analysis of hyperdimensional feature spaces, to evaluate their effectiveness in classifying complex forest areas in a multisource framework. The first technique is a parametric regularized GML classifier that applies the leave-one-out-covariance (LOOC) procedure [48] to the estimation of the statistics of the classifier. The second technique is a distribution-free machine learning classifier based on the SVM [49], [50]. The main motivations for this choice are: 1) GML-LOOC and SVM have been widely used in previous studies on classification of hyperspectral data (e.g., [48]–[50]), proving their effectiveness in hyperdimensional feature spaces; 2) both techniques are intrinsically able to solve ill-posed classification problems, in which the ratio between the number of available training samples and the number of features is relatively small (this is a typical situation with hyperspectral data); and 3) despite the aforementioned common properties, GML-LOOC and SVM represent a good sampling of two different categories of classification algorithms. GML-LOOC is a parametric classifier based on the Gaussian model for the approximation of the class distributions. It represents an effective version of the widely used standard ML classifier for the analysis of hyperspectral data. The SVM classifier is a distribution-free complex classifier, which is based on machine learning and, thus, on a completely different theoretical background with respect to GML-LOOC. SVM proved to be very effective for classification of hyperspectral data (e.g., [49] and [50]).

In the following, we briefly recall the main properties of these classifiers.

#### A. GML-LOOC

This algorithm belongs to the family of parametric techniques and is based on the GML classifier. The standard GML procedure is effective when the ratio between the number of training samples and the dimension of the feature space is relatively high, but its performance degrades when this ratio decreases (Hughes phenomenon [43]). In particular, when the number of training samples is smaller than the number of features, the covariance matrix used in the decision rule become singular, and thus, the GML cannot be used. To avoid this problem, several algorithms have been developed for the estimation of a nonsingular covariance matrix (e.g., [48]–[54]). In our study, we chose the algorithm proposed in [48], called LOOC algorithm. In the following, we give more details on this classifier.

Let  $\mathbf{x}_p$  be the  $p$ th pattern to be classified, and  $\mu_i$  and  $\Sigma_i$  (with  $i = 1, \dots, K$ ) be the mean value and the covariance matrix of the  $i$ th investigated class, respectively. The decision rule is as follows:

$$\mathbf{x}_p \in \omega_j \Leftrightarrow d_j(\mathbf{x}_p) > d_i(\mathbf{x}_p) \quad \forall i \neq j \quad (1)$$

where  $d_i(\mathbf{x}_p)$  is computed as

$$d_i(\mathbf{x}_p) = (\mathbf{x}_p - \mu_i)^t \Sigma_i^{-1} (\mathbf{x}_p - \mu_i) + \ln |\Sigma_i|. \quad (2)$$

Usually, the true values of the mean vectors and of the covariance matrices are not known, and they should be estimated from the training samples. When a reduced number of samples is available, the covariance matrices can be replaced with the common covariance matrix, defined as  $\mathbf{S} = (1/K) \sum_{i=1}^K \Sigma_i$  [48]. The LOOC algorithm proposes a more refined way to estimate the covariance matrices for classes characterized by a reduced number of training samples. In particular, the covariance matrix  $\Sigma_i^{\text{LOOC}}$  of the  $i$ th class is estimated as follows:

$$\Sigma_i^{\text{LOOC}}(\alpha_i) = \begin{cases} (1 - \alpha_i) \text{diag}(\Sigma_i) + \alpha_i \Sigma_i, & 0 \leq \alpha_i \leq 1 \\ (2 - \alpha_i) \Sigma_i + (\alpha_i - 1) \mathbf{S}, & 1 < \alpha_i \leq 2 \\ (3 - \alpha_i) \mathbf{S} + (\alpha_i - 2) \text{diag}(\mathbf{S}), & 2 < \alpha_i \leq 3 \end{cases} \quad (3)$$

where  $\alpha_i$  is a mixing parameter, whose value is selected according to the following procedure: 1) removing one sample; 2) computing the mean and covariance from the remaining samples; and 3) computing the likelihood of the sample which was left out, given the mean and covariance estimates. Each sample is removed in turn, and the average log likelihood is computed. The value that maximizes the average log likelihood is selected [55]. This implementation has proved to be particularly effective in hyperspectral data classification.<sup>1</sup>

It is worth noting that since this classifier models the class distributions according to a Gaussian function, its application to multisensor data implies a Gaussian approximation of the

distribution of classes on the stacked feature vector. This approximation is reasonable from an application viewpoint, but it is not rigorous from a theoretical perspective.

#### B. SVM

SVMs are distribution-free classifiers that overcome the aforementioned approximation of the GML-LOOC classifier. Developed by Vapnik [56], SVM classifiers have undergone great development in the last ten years and have been successfully applied to several remote sensing problems (e.g., [49] and [50]). Their success is justified from four main properties: 1) their relatively high classification accuracy and very good generalization capability with respect to other classifiers; 2) the limited effort required for architecture design and training phase if compared to other machine learning algorithms (such as multilayer perceptron neural networks); 3) the convexity of the cost function that always finds the optimum solution; and 4) their effectiveness in ill-posed classification problems (problems with a low ratio between number of training samples and number of features) [50]. In the following, we briefly relate the main concepts and the mathematical formulation of SVMs.<sup>2</sup>

Let us consider a binary classification problem. Let us assume that the training set consists of  $Q$  vectors  $\mathbf{x}_p \in \mathbb{R}^q$ , with the corresponding target  $\mathbf{y}_p \in \{-1; +1\}$ , where “+1” and “−1” denote the labels of the considered classes. The nonlinear SVM approach consists of mapping the data into a higher dimensional feature space, i.e.,  $\Phi(\mathbf{x}_p) \in \mathbb{R}^{q'}$  ( $q' \gg q$ ), where it looks for a separation between the two classes by means of an optimal hyperplane defined by a weight vector  $\mathbf{w} \in \mathbb{R}^{q'}$  and a bias  $b \in \mathbb{R}$ . In particular,  $\mathbf{w}$  is a vector orthogonal to the separating hyperplane,  $b$  is a scalar value such that the ratio  $b/\|\mathbf{w}\|$  represents the distance of the hyperplane from the origin, and the function  $\Phi$  represents a nonlinear transformation. The membership decision rule is defined according to  $\text{sign}[f(\mathbf{x})]$ , where  $f(\mathbf{x})$  represents the discriminant function associated with the hyperplane and is written as

$$f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b. \quad (4)$$

The optimal hyperplane is the one that minimizes a cost function, which expresses a combination of two criteria: margin maximization and error minimization. It is defined as

$$\Psi(\mathbf{w}, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{p=1}^Q \xi_p. \quad (5)$$

This cost function minimization is subject to the following constraints:

$$\mathbf{y}_p \cdot (\mathbf{w} \cdot \mathbf{x}_p + b) \geq 1 - \xi_p \quad \forall p = 1, \dots, Q \quad (6)$$

$$\xi_p \geq 0 \quad \forall p = 1, \dots, Q \quad (7)$$

where  $\xi_p$  are the so-called *slack variables* and are defined as follows:

$$\xi((\mathbf{x}_p, \mathbf{y}_p), (\mathbf{w}, b)) = \xi_p = \max(0, 1 - \mathbf{y}_p (\mathbf{w} \cdot \Phi(\mathbf{x}_p) + b)) \quad (8)$$

<sup>1</sup>In this paper, we used the implementation contained in the MultiSpec software [55].

<sup>2</sup>We used our own implementation of SVM, which is based on the sequential minimal optimization procedure.



and they are introduced to take into account nonseparable data. The constant  $C$  represents a regularization parameter that controls the shape of the discriminant function, and consequently, the decision boundary when data are nonseparable. The above optimization problem can be reformulated through a Lagrange functional for which the Lagrange multipliers can be found by means of a dual optimization leading to a quadratic programming solution [56]. The final result is a discriminant function conveniently expressed as a function of the data in the original (lower) dimensional feature space, i.e.,

$$f(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i \Phi(\mathbf{x}_i, \mathbf{x}) + b \quad (9)$$

where  $\Phi(\cdot, \cdot)$  is a kernel function, and  $S$  is the subset of training samples corresponding to the nonzero Lagrange multipliers. It is worth noting that the Lagrange multipliers  $\alpha_i$  effectively weight each training sample according to its importance in determining the discriminant function. The training samples associated with nonzero weights are termed *support vectors* [50]. In particular, the support vectors, where  $\alpha_i = C$ , are referred to as *bound support vector*, and support vectors with  $0 < \alpha_i < C$  are called *nonbound support vectors*. The kernel  $\Phi(\cdot, \cdot)$  must satisfy the condition of Mercer's theorem so that it corresponds to some type of inner product in the transformed (higher) dimensional feature space [56].

The SVM classifier was developed to solve binary classification problems, but it can be easily extended to multiclass problems. The two main strategies used for  $K$  class problems are as follows. 1) *One-Against-One*—the  $K$ -class problem is decomposed into  $K(K-1)/2$  binary problems, each focused on the recognition of a pair of classes. A generic pattern is associated with the class that receives the majority of the votes from the ensemble of binary classifiers. 2) *One-Against-All*—the  $K$ -class problem is decomposed into  $K$  binary problems, each focused on the recognition of one class against all the others. The “winner-takes-all” rule is used for the final decision, i.e., the winning class is the one corresponding to the SVM with the highest output (discriminant function value). We refer the reader to [50] for greater details on SVM classifiers and on the related multiclass strategies.

## V. EXPERIMENTAL ANALYSIS AND DISCUSSION

### A. Experimental Design

To assess the effectiveness of the proposed system and to achieve the goals of this paper, we defined three different experiments: 1) analysis of the importance of the joint use of hyperspectral images and the first LIDAR return on the classification of complex forest areas; 2) analysis on the usefulness of multiple LIDAR returns and of the different information contained in elevation and intensity channels; and 3) analysis on the generalization capability of the proposed system.

For the first two experiments, we carried out the learning of the classifier (with the model selection) and the accuracy assessment according to a  $k$ -fold cross-validation procedure. This allowed us to analyze, from a rigorous statistical perspective, the potential of the proposed system, and of the hyperspectral

and LIDAR sensors, in the considered scenario. We randomly divided the available ground truth data into five subsets, and we then adopted a fivefold cross-validation procedure, with training samples (pixels) distributed all over the scene. The samples of ground truth data available were used as follows: 20% in the training set (about 3300 samples) and 80% in the test set (about 13 500 samples). It is worth noting that the use of only 20% of the ground truth samples for learning tends to result in minority classes with very few training samples. However, this choice is reasonable as it represents a typical condition of real operational applications.

With regard to the last experiment, we defined the training and the test sets by considering samples from different spatially disjoint areas to analyze the generalization capability of the system with respect to the variability and the nonstationary behavior of the spectral signatures of the classes. In further detail, for this experiment, the training and test samples were selected purposefully avoiding that they share pixels belonging to the same tree crown (i.e., all the pixels of a tree crown are completely included in only one of the two sets).

The performances of the system were assessed by using error matrices. We derived the overall kappa coefficients from these matrices, as described by Congalton and Green [57], and analyzed the statistic significance of results according to the Zeta test [57].

In our experiments, we used also the  $k$ -Nearest Neighbor ( $k$ -NN) classifier to compare the accuracy provided by the advanced classifiers included in the proposed system with a simple distribution-free classification technique. For the model selection of the SVM classifier, we chose a Gaussian kernel function and applied a grid search strategy in a range between 50 and 240 for  $C$  and in a range between 1 and 1000 for  $\gamma$ . For the  $k$ -NN classifier, the value of  $k$  varied from 1 to 29.

### B. Experiment 1: Analysis of the Effectiveness of the Proposed Multisensor Classification System

Let us consider the noise-reduced hyperspectral data, as well as the intensity and the corrected elevation of the first LIDAR return. In this experiment, we analyze the effectiveness of the first LIDAR return channels, at first, considering only the global kappa accuracy, and then analyzing in greater detail the class-by-class accuracies. Experiments were conducted with three classifiers: SVM, GML-LOOC, and  $k$ -Nearest Neighbor ( $k$ -NN). We carried out different trials using 126, 40, and 25 spectral channels derived according to the feature selection algorithm. Fig. 3 shows the behavior of the average Jeffreys–Matusita distance versus the number of hyperspectral channels selected with the SFFS search strategy. It is worth noting that the typical trend of this distance, which reaches saturation when the number of features used do not change the separability among information classes. We reached saturation with about 25 features but, in this experiment and in the following, we also analyzed what occurred with 40 hyperspectral features. This was done for consideration of some margin on the minimum number of input channels derived from the feature selection phase. This is reasonable to better consider also the accuracy of minority classes that less affect the behavior of the average Jeffreys–Matusita distance.

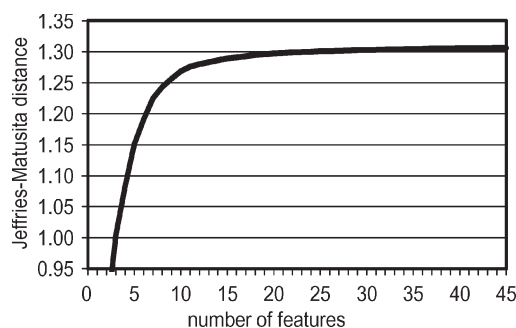


Fig. 3. Behavior of the average Jeffries–Matusita distance versus the number of selected features.

TABLE II  
KAPPA ACCURACIES OBTAINED ON THE TEST SET WITH DIFFERENT SPECTRAL FEATURES AND CLASSIFIERS

FEATURES	KAPPA ACCURACY		
	SVM	GML-LOOC	k-NN
25 HYPERSPECTRAL BANDS	0.872	0.778	0.649
40 HYPERSPECTRAL BANDS	0.879	0.782	0.666
126 HYPERSPECTRAL BANDS	0.881	0.823	0.676

Table II illustrates the kappa accuracies obtained with different classifiers when varying the features used. From the analysis of these accuracies, we can infer some important points. For the SVM classifier, the accuracies obtained are particularly high considering the number of classes (23) and the number of training samples per class (as shown in Table I, for some classes in the training phase, we have only ten samples). In particular, it is possible to observe that the SVM classifier always provided significantly higher accuracy than both the  $k$ -NN and the GML-LOOC techniques. In greater detail, due to its strongly nonlinear properties, the SVM classifier obtained with 25 features a kappa accuracy that is higher than that obtained by the GML-LOOC technique with all the 126 channels. These results confirm the superior performances of the SVM technique, which also involves an intrinsically better generalization ability. The higher potentialities of the SVM classifier can be explained by the fact that it is a distribution-free technique that does not approximate the distribution of classes with any predefined statistical model (the GML-LOOC assume Gaussian approximation) but models the decision boundary on the basis of the available training data. This results in the capability to model also strongly nonlinear decision boundaries. Another important issue to note is that both SVM and GML-LOOC do not seem significantly affected by the Hughes phenomenon, since the classification accuracies increase with the increase of the number of features. In the case of  $k$ -NN, the kappa accuracies for all the experiments were much smaller than those obtained by the other classifiers. This confirms that  $k$ -NN is not able to manage hyperdimensional feature spaces. This is particularly true when classes with very few training samples are considered. The small kappa accuracies also illustrate the importance of using specific classifiers that exhibit a high generalization ability.

Let us now analyze the effect of the first LIDAR return channels on the classification accuracy. Table III shows the

TABLE III  
KAPPA ACCURACIES OBTAINED ON THE TEST SET WITH DIFFERENT SPECTRAL AND LIDAR FEATURES AND CLASSIFIERS

FEATURES	KAPPA ACCURACY		
	SVM	GML-LOOC	k-NN
25 HYPERSPECTRAL BANDS + ELEVATION AND INTENSITY OF THE FIRST LIDAR RETURN	0.885	0.809	0.698
40 HYPERSPECTRAL BANDS + ELEVATION AND INTENSITY OF THE FIRST LIDAR RETURN	0.890	0.809	0.714
126 HYPERSPECTRAL BANDS + ELEVATION AND INTENSITY OF THE FIRST LIDAR RETURN	0.892	0.840	0.714

accuracies obtained adding to different spectral feature subsets the elevation and intensity channels of the first LIDAR return. Comparing the results in Tables II and III reveals an increase in kappa accuracy from 1% to 4%, which is less relevant for the SVM and GML-LOOC classifiers and more significant for the  $k$ -NN technique (which, however, does not obtain acceptable accuracies). From these results, it seems that LIDAR channels provide relatively sparse information for discriminating between tree species. However, if we analyze the class-by-class accuracies, with and without LIDAR channels, the conclusions are quite different. Table IV shows class-by-class accuracies obtained with the SVM classifier, adding the two LIDAR channels to different spectral band subsets (25 and 40). From the analysis of the table, we observe that, in general, very high accuracies were reached for very similar tree species, including *Quercus cerris*, *Quercus robur*, and *Quercus rubra*. Concerning the role of LIDAR channels, we have different classification behaviors varying the number of spectral bands used. When 40 bands were used, an increase in the classification accuracy occurs for classes characterized by a low height. In particular, for four of the classes, the increment was higher than 5%. The increase in the classification accuracy becomes more relevant when reducing the number of hyperspectral bands used. With 25 hyperspectral channels, the accuracy increased by more than 10% for two classes and more than 5% for seven classes. For example, *Acer negundo* increased in accuracy by 13.56%, adding LIDAR bands to 25 hyperspectral channels. Analyzing the confusion matrices, this class increases its separability with respect to *Carpinus betulus*, *Platanus hybrida*, and *Quercus robur* (which are characterized by a very different mean height with respect to *Acer negundo*). It is possible to draw similar conclusions also for others classes that have relevant increase in the classification accuracy. It is worth noting that the classes that significantly increase their accuracy by introducing LIDAR features are the underrepresented classes. This is the motivation for the relatively small impact of this improvement on the overall classification accuracy.

The Zeta test [57] was computed between kappa accuracies obtained with 40 hyperspectral bands, with and without LIDAR channels, using SVM and GML-LOOC classifiers. All the differences in accuracy were statistically significant at 95% of the confidence interval.



TABLE IV  
CLASS-BY-CLASS ACCURACY OBTAINED BY THE SVM CLASSIFIER WITH LIDAR AND  
WITHOUT LIDAR WITH A DIFFERENT NUMBER OF HYPERSPECTRAL CHANNELS

CLASS NAME	25 HYPERSPECTRAL FEATURES			40 HYPERSPECTRAL FEATURES		
	ACCURACY (%)		$\Delta$ (%)	ACCURACY (%)		$\Delta$ (%)
	LIDAR	NO LIDAR		LIDAR	NO LIDAR	
<i>Acer campestris</i>	76.76	70.29	6.47	75.59	71.03	4.56
<i>Acer negundo</i>	85.45	71.89	13.56	87.53	80.19	7.34
<i>Alnus glutinosa</i>	90.63	91.12	-0.49	91.02	90.73	0.30
<i>Carpinus betulus</i>	92.39	91.54	0.85	92.31	91.68	0.63
<i>Corylus avellana</i>	45.19	38.75	6.44	45.62	38.34	7.28
<i>Fraxinus angustifolia</i>	90.34	90.82	-0.48	90.25	89.83	0.41
Grassland	100.00	98.64	1.36	100.00	99.09	0.91
<i>Juglans nigra</i>	86.57	84.32	2.26	88.25	85.19	3.06
<i>Juglans regia</i>	89.67	90.40	-0.73	90.67	90.69	-0.02
<i>Morus sp.</i>	87.34	75.92	11.42	87.34	78.67	8.67
<i>Platanus hybrida</i>	89.99	90.23	-0.24	90.20	90.93	-0.73
<i>Populus canescens</i>	87.29	84.84	2.46	89.04	87.19	1.84
<i>Populus hybrida</i>	89.10	85.66	3.44	89.92	87.31	2.61
<i>Prunus avium</i>	79.79	72.89	6.90	80.65	76.53	4.12
<i>Quercus cerris</i>	92.87	93.07	-0.19	93.35	92.80	0.54
<i>Quercus robur</i>	86.49	86.91	-0.41	88.25	88.24	0.01
<i>Quercus rubra</i>	92.22	93.42	-1.21	91.93	93.36	-1.43
<i>Robinia pseudacacia</i>	88.54	85.05	3.50	88.22	84.90	3.32
<i>Rubus</i>	93.72	86.46	7.26	93.19	87.75	5.45
Shadows	98.02	97.93	0.09	98.28	97.84	0.43
Snags	86.34	85.98	0.37	85.61	85.85	-0.24
<i>Tilia cordata</i>	89.69	84.07	5.62	89.30	86.04	3.26
<i>Ulmus minor</i>	70.22	65.81	4.40	71.65	69.17	2.48

From these results, it is possible to conclude that the first LIDAR return can be very useful in the discrimination of specific tree species. In addition, if LIDAR channels are available, it is convenient to use a reduced number of spectral channels and to add to these channels the LIDAR information. In this perspective, on the one hand, the LIDAR channels compensate the minor lost of information due to the reduced number of spectral channels; on the other hand, the resulting smaller number of features allows both a reduction of computation time and an increase in the generalization capability of the system.

### C. Experiment 2: Detailed Analysis of the Complementary Information Contained in LIDAR Returns

In the first part of this experiment, we considered 40 hyperspectral bands and the two channels of the first LIDAR return. From the results of the previous experiment, it is clear that these channels are useful in classification of complex forest areas, particularly for discriminating between specific tree species. The next step is to understand the amount of information present in each channel (i.e., elevation and intensity).

Table V presents the kappa accuracies obtained with the SVM, with: 1) 40 hyperspectral bands; 2) 40 hyperspectral bands plus elevation and intensity of the first LIDAR return; 3) 40 hyperspectral bands plus elevation of the first LIDAR return; and 4) 40 hyperspectral bands plus intensity of the first LIDAR return. As it is clear from the table, the increase in the classification accuracy obtained with LIDAR data is mainly due to the elevation channel, whereas the intensity channel does not give any relevant information for the classification of the considered forest area (it slightly decreases the overall kappa accuracy).

TABLE V  
KAPPA ACCURACIES OBTAINED WITH SVM VARYING THE  
FIRST LIDAR RETURN CHANNELS USED

FEATURES USED	KAPPA ACCURACY
40 HYPERSPECTRAL FEATURES	0.879
40 HYPERSPECTRAL FEATURES + ELEVATION AND INTENSITY OF THE FIRST LIDAR RETURN	0.890
40 HYPERSPECTRAL FEATURES + ELEVATION OF THE FIRST LIDAR RETURN	0.888
40 HYPERSPECTRAL FEATURES + INTENSITY OF THE FIRST LIDAR RETURN	0.876

TABLE VI  
KAPPA ACCURACIES OBTAINED WITH SVM VARYING THE  
NUMBER OF LIDAR RETURNS JOINTLY USED WITH  
THE 40 HYPERSPECTRAL FEATURES SELECTED

LIDAR FEATURES USED	KAPPA ACCURACY
1 <sup>ST</sup> RETURN CHANNELS	0.890
1 <sup>ST</sup> + 2 <sup>ND</sup> RETURN CHANNELS	0.878
1 <sup>ST</sup> + 2 <sup>ND</sup> + 3 <sup>RD</sup> RETURN CHANNELS	0.872
1 <sup>ST</sup> + 2 <sup>ND</sup> + 3 <sup>RD</sup> + 4 <sup>TH</sup> RETURN CHANNELS	0.872

The second part of this experiment was focused on multiple LIDAR returns available in the data set used for this paper. To analyze the information contained in these channels for the classification process, we carried out a series of trials incrementally adding the first, the second, the third, and the fourth return channels (elevation and intensity) to the 40 hyperspectral bands.

Table VI shows the overall kappa accuracies obtained in these trials. From these results, it seems that returns different from the first do not increase kappa accuracy. On the contrary, they result in a slight decrease of the accuracy with respect to

TABLE VII  
KAPPA ACCURACIES OBTAINED ON TEST SET WITH DIFFERENT  
CLASSIFIERS USING DISJOINT TRAINING AND TEST SETS

FEATURES USED	KAPPA ACCURACY		
	SVM	GML-LOOC	k-NN
40 HYPERSPECTRAL BANDS	0.691	0.629	0.468
40 HYPERSPECTRAL BANDS + ELEVATION AND INTENSITY OF THE FIRST LIDAR RETURN	0.717	0.658	0.484

that yielded using hyperspectral features plus the first LIDAR return channels. These results depend on the properties of available multiple LIDAR returns. As described in Section II, the number of pulses is different for each return, and, in particular, it decreases by increasing the return number. This can be explained by the fact that the analyzed area is characterized by a very dense tree crown coverage that precludes the generation of secondary returns in many portions of the scene. For this reason, during the rasterization phase of the LIDAR data, not all the pixels were associated with a value. In particular, for returns 3 and 4, we have many pixels with no data points. This introduces a noise in the classification process, thus, balancing possible advantages in the characterization of the canopy of different species. In general, we expect that this issue should be better investigated using data with a higher number of representative samples from multiple returns.

#### D. Experiment 3: Generalization Capability of the System

As described in Section V-A, with this last experiment, we simulate a borderline case, in which training and test samples are as disjointed as possible. This allows us to verify the behavior of the proposed system when test samples belong to a significantly different area from the one considered for the training of the system. It is worth noting that, in this specific case, we have some classes with a reduced number of ground truth samples; this means that only few (three or four) trees in the whole scene were available for these classes. Thus, the exclusive assignment of a tree to the training or the test set makes ground truth for minority classes unrepresentative of the variability of the spectral signature over the scene, resulting in a very difficult classification problem.

In Table VII, one can see that the kappa accuracies decrease with respect to the previous experiments for all the three classifiers considered. On the contrary, the differences in accuracies between the two subsets of features (with and without LIDAR) remain almost the same. The kappa accuracy of SVM was still significantly higher than those provided by other classifiers (i.e., 0.717 versus 0.658 and 0.484), but there was a large decrease with respect to those yielded in previous experiments. In addition, the GML-LOOC significantly decreased the kappa accuracy. However, as expected, the more relevant degradation was associated with the  $k$ -NN classifier, that resulted in a kappa accuracy lower than 50%. Analyzing the SVM class-by-class accuracies, we observed that for some dominant classes, including *Carpinus betulus*, *Juglans regia*, *Platanus hybrida*, and *Quercus rubra*, the accuracies are still on the order of 85%–90% (88.41%, 87.43%, 84.35%, and

93.47%, respectively), whereas, for the minority classes, we have a dramatic decrease of accuracies. For example, for *Acer campestre*, *Populus hybrida*, *Prunus avium*, and *Ulmus minor*, the accuracies were lower than 50% (42.86%, 30.56%, 32.39%, and 24.31%, respectively). These results were expected in this very critical scenario (see [58]) that should be addressed by using semisupervised classification techniques (like semisupervised SVM [59]) particularly developed for strongly ill-posed problems.

The differences in kappa accuracy between trials with and without LIDAR channels, with SVM and GML-LOOC classifiers, were also tested with the  $Z$ -test [57]. All the differences resulted to be statistically significant at 95% of the confidence interval.

## VI. CONCLUSION

In this paper, we investigated the joint use of hyperspectral and LIDAR remote sensing data for the classification of complex forest areas. We analyzed this issue by proposing a novel classification system, based on different possible classifiers that were able to properly integrate multisensor information. From an analysis of the results of all the experiments carried out using the proposed system, we can conclude that, in general, it provided high accuracies, managing in an effective way the complementary information contained in hyperspectral and LIDAR data. In greater detail, we verified the following.

- 1) The presented system is very effective for classifying hyperspectral and LIDAR data, providing high accuracy on almost all the considered forest classes (it yielded accuracies of over 90% for certain classes).
- 2) The distribution-free SVM classifier provided much higher accuracies than the other classifiers investigated. The parametric GML-LOOC, even if less effective than SVM, yielded acceptable accuracies, whereas the  $k$ -NN technique (used for comparison) was unsuitable for the solution of hyperdimensional problems.
- 3) The elevation channel of the first LIDAR return data played the most important role for increasing the discriminability (and, thus, the accuracy) of the forest classes by having similar spectral signatures. This was due to the different average elevation of some forest classes.
- 4) LIDAR returns that are different from the first return do not seem capable of improving the kappa accuracy when used jointly with hyperspectral channels. However, this issue should be better analyzed on other data sets by considering a more complex feature extraction phase.
- 5) In critical cases, with a large difference between training and test samples, the system based on the SVM classifier should provide an acceptable accuracy. However, in this extreme case, the performances were degraded significantly, and it is recommended to use specific classification techniques developed for ill-posed problems (e.g., semisupervised [59]).

As a final remark, it is important to observe that the proposed system and study seem particularly relevant when considering that, in several forest areas, both hyperspectral and LIDAR data are acquired for species classification and parameter estimation, respectively. In these situations, it is important to properly

integrate LIDAR data in the classification process because: 1) the use of hyperspectral and LIDAR data increases the separability of tree species having similar spectral signatures but different height and 2) the introduction of the first LIDAR return elevation channel produces, with a limited number of spectral features, accuracies similar to those yielded with a significantly higher number of features. This results in a lower computational time and in an increase of the generalization capability of the system.

In terms of future developments of this paper, we are planning to: 1) introduce, in the classification phase, semisupervised classifiers to increase the generalization ability of the system and improve the modeling of the nonstationarity of the spectral signatures of classes in the scene [58] and 2) jointly exploit hyperspectral images and LIDAR data for the estimation of biophysical forest parameters (e.g., biomass and structure).

#### ACKNOWLEDGMENT

The authors would like to thank Dr. F. Mason (Centro Nazionale per lo Studio e la Conservazione della Biodiversità Forestale—Bosco della Fontana—Verona) for providing the hyperspectral and LIDAR data used in this paper.

#### REFERENCES

- [1] A. A. Abuelgasim, R. A. Fernandes, and S. G. Leblanc, "Evaluation of national and global LAI products derived from optical remote sensing instruments over Canada," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 7, pp. 1872–1884, Jul. 2006.
- [2] F. Maselli and M. Chiesi, "Evaluation of statistical methods to estimate forest volume in a Mediterranean region," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2239–2250, Aug. 2006.
- [3] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2563–2586, Sep. 2006.
- [4] C. Song, T. A. Schroeder, and W. B. Cohen, "Predicting temperate conifer forest successional stage distributions with multitemporal Landsat Thematic Mapper imagery," *Remote Sens. Environ.*, vol. 106, no. 2, pp. 228–237, Jan. 2007.
- [5] F. Sedano, P. Gong, and M. Ferrao, "Land cover assessment with MODIS imagery in southern African Miombo ecosystems," *Remote Sens. Environ.*, vol. 98, no. 4, pp. 429–441, Oct. 2005.
- [6] D. G. Goodenough, A. Dyk, K. O. Niemann, J. S. Pearlman, H. Chen, T. Han, M. Murdoch, and C. West, "Processing Hyperion and ALI for forest classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1321–1331, Jun. 2003.
- [7] N. Kosaka, T. Akiyama, B. Tsai, and T. Kojima, "Forest type classification using data fusion of multispectral and panchromatic high-resolution satellite images," in *Proc. IGARSS*, Jul. 25–29, 2005, vol. 4, pp. 2980–2983.
- [8] L. Wang, W. P. Sousa, P. Gong, and G. S. Biging, "Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama," *Remote Sens. Environ.*, vol. 91, no. 3/4, pp. 432–440, Jun. 2004.
- [9] D. M. Gates, H. J. Keegan, J. C. Schleiter, and V. R. Weidner, "Spectral properties of plants," *Appl. Opt.*, vol. 4, no. 1, pp. 11–20, Jan. 1965.
- [10] E. B. Knipling, "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation," *Remote Sens. Environ.*, vol. 1, no. 3, pp. 155–159, 1970.
- [11] P. J. Zarco-Tejada, J. R. Millera, J. Harrona, B. Hub, T. L. Noland, N. Goele, G. H. Mohammed, and P. Sampson, "Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies," *Remote Sens. Environ.*, vol. 89, no. 2, pp. 189–199, Jan. 2004.
- [12] M. E. Martin, S. D. Newman, J. D. Aber, and R. G. Congalton, "Determining forest species composition using high spectral resolution remote sensing data," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 249–254, Sep. 1998.
- [13] M. L. Clark, D. A. Roberts, and D. B. Clark, "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales," *Remote Sens. Environ.*, vol. 96, no. 3/4, pp. 375–398, Jun. 2005.
- [14] D. G. Leckie, S. Tinis, T. Nelson, C. Burnett, F. A. Gougeon, E. Cloney, and D. Paradine, "Issues in species classification of trees in old growth conifer stands," *Can. J. Remote Sens.*, vol. 31, no. 2, pp. 175–190, 2005.
- [15] M. Watanabe, M. Shimada, A. Rosenqvist, T. Tadono, M. Matsuoka, S. A. Romshoo, K. Ohta, R. Furuta, K. Nakamura, and T. Moriyama, "Forest structure dependency of the relation between L-band  $\sigma^0$  and biophysical parameters," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3154–3165, Nov. 2006.
- [16] S. Saatchi, K. Halligan, D. G. Despain, and R. L. Crabtree, "Estimation of forest fuel load from radar remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1726–1740, Jun. 2007.
- [17] H. Balzer, L. Skinner, A. Luckman, and R. Brooke, "Estimation of tree growth in a conifer plantation over 19 years from multi-satellite L-band SAR," *Remote Sens. Environ.*, vol. 84, no. 2, pp. 184–191, Feb. 2003.
- [18] L. Bruzzone, M. Marconcini, U. Wegmuller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, Jun. 2004.
- [19] J. S. Lee, K. P. Papathanassiou, I. Hajnsek, T. Mette, M. R. Grunes, T. Ainsworth, and L. Ferro-Famil, "Applying polarimetric SAR interferometric data for forest classification," in *Proc. IGARSS*, Jul. 25–29, 2005, vol. 7, pp. 4848–4851.
- [20] K. J. Ranson, G. Sun, V. I. Kharuk, and K. Kovacs, "Characterization of forests in Western Sayani mountains, Siberia from SIR-C SAR data," *Remote Sens. Environ.*, vol. 75, no. 2, pp. 188–200, Feb. 2001.
- [21] S. S. Saatchi and E. Rignot, "Classification of boreal forest cover types using SAR images," *Remote Sens. Environ.*, vol. 60, no. 3, pp. 270–281, Jun. 1997.
- [22] M. Maltamo, K. Eerikainen, J. Pitkanen, J. Hyypä, and M. Vehmas, "Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 319–330, Apr. 2004.
- [23] B. Koetz, F. Morsdorf, G. Sun, K. J. Ranson, K. Itten, and B. Allgöwer, "Inversion of a LIDAR waveform model for forest biophysical parameter estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 49–53, Jan. 2006.
- [24] N. Skowronski, K. Clark, R. Nelson, J. Hom, and M. Patterson, "Remotely sensed measurements of forest structure and fuel loads in the pinelands of New Jersey," *Remote Sens. Environ.*, vol. 108, no. 2, pp. 123–129, May 2007.
- [25] H. E. Andersen, R. J. McGaughey, and S. E. Reutebuch, "Estimating forest canopy fuel parameters using LIDAR data," *Remote Sens. Environ.*, vol. 94, no. 4, pp. 441–449, Feb. 2005.
- [26] R. Brennan and T. L. Webster, "Object-oriented land cover classification of LIDAR-derived surfaces," *Can. J. Remote Sens.*, vol. 32, no. 2, pp. 162–172, 2006.
- [27] A. P. Charaniya, R. Manduchi, and S. K. Lodha, "Supervised parametric classification of aerial LiDAR data," in *Proc. IEEE Comput. Soc. Conf. CVPRW*, 2004, p. 30.
- [28] J. Holmgren and Å. Persson, "Identifying species of individual trees using airborne laser scanner," *Remote Sens. Environ.*, vol. 90, no. 4, pp. 415–423, Apr. 2004.
- [29] M. Moghaddam, J. L. Dungan, and S. Acker, "Forest variable estimation from fusion of SAR and multispectral optical data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2176–2187, Oct. 2002.
- [30] R. M. Lucas, N. Cronin, M. Moghaddam, A. Lee, J. Armston, P. Bunting, and C. Witte, "Integration of radar and Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 388–406, Feb. 2006.
- [31] A. T. Hudak, M. A. Lefsky, W. B. Cohen, and M. Berterretche, "Integration of LIDAR and Landsat ETM+ data for estimating and mapping forest canopy height," *Remote Sens. Environ.*, vol. 82, no. 2/3, pp. 397–416, Oct. 2002.
- [32] G. A. Blackburn, "Remote sensing of forest pigments using airborne imaging spectrometer and LIDAR imagery," *Remote Sens. Environ.*, vol. 82, no. 2/3, pp. 311–321, Oct. 2002.
- [33] P. Hyde, R. Nelson, D. Kimes, and E. Levine, "Exploring LiDAR–RaDAR synergy—Predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR," *Remote Sens. Environ.*, vol. 106, no. 1, pp. 28–38, Jan. 2007.
- [34] P. Hyde, R. Dubayah, W. Walker, J. B. Blair, M. Hofton, and C. Hunsaker, "Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy," *Remote Sens. Environ.*, vol. 102, no. 1/2, pp. 63–73, May 2006.

- [35] E. Simental, D. J. Ragsdale, E. Bosch, R. Dodge, Jr., and R. Pazak, "Hyperspectral dimension reduction and elevation data for supervised image classification," in *Proc. 14th ASPRS Conf.*, Anchorage, AK, May 3–9, 2003.
- [36] D. Lemp and U. Weidner, "Improvements of roof surface classification using hyperspectral and laser scanning data," in *Proc. ISPRS Joint Conf.: 3rd Int. Symp. Remote Sens. Data Fusion Over Urban Areas (URBAN)*, 5th Int. Symp. Remote Sens. Urban Areas (URS), Tempe, AZ, Mar. 14–16, 2005.
- [37] J. T. Mundt, D. R. Streutker, and N. F. Glenn, "Mapping sagebrush distribution using fusion of hyperspectral and LIDAR classifications," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 1, pp. 47–54, Jan. 2006.
- [38] R. Sugumaran and M. Voss, "Object-oriented classification of LIDAR-fused hyperspectral imagery for tree species identification in an urban environment," in *Proc. Urban Remote Sens. Joint Event*, Paris, France, Apr. 11–13, 2007, pp. 1–6.
- [39] E. M. Perry, H. P. Foote, G. M. Petrie, K. L. Steinmaus, D. E. Irwin, and A. J. Stephan, "Exploitation of hyperspectral imagery and Lidar for landuse classification," in *Proc. IGARSS*, Seattle, WA, 1998, pp. 1013–1015.
- [40] D. Yuan and C. D. Elvidge, "Comparison of relative radiometric normalization techniques," *ISPRS J. Photogramm. Remote Sens.*, vol. 51, no. 3, pp. 117–126, Jun. 1996.
- [41] Y. Du, J. Cihlar, J. Beaubien, and R. Latifovic, "Radiometric normalization, compositing, and quality control for satellite high resolution image mosaics over large areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 623–634, Mar. 2001.
- [42] P. Hsieh and D. Landgrebe, "Lowpass filter for increasing class separability," in *Proc. IGARSS*, 1998, pp. 2691–2693.
- [43] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*. New York: Springer-Verlag, 1999.
- [44] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
- [45] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.
- [46] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, Jul. 2001.
- [47] L. Bruzzone, F. Roli, and S. B. Serpico, "An extension of the Jeffreys–Matusita distance to multiclass cases for feature selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 6, pp. 1318–1321, Nov. 1995.
- [48] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 7, pp. 763–767, Jul. 1996.
- [49] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
- [50] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [51] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, Mar. 1989.
- [52] S. Marks and O. J. Dunn, "Discriminant functions when the covariance matrices are unequal," *J. Amer. Stat. Assoc.*, vol. 69, no. 346, pp. 555–559, Jun. 1974.
- [53] S. P. Lin and M. D. Perlman, "A Monte Carlo comparison of four estimators of a covariance matrix," in *Proc. 6th Int. Symp. Multivariate Anal.*, 1985, pp. 411–429.
- [54] P. W. Wahl and R. A. Kronmull, "Discriminant functions when covariances are unequal and sample sizes are moderate," *Biometrics*, vol. 33, no. 3, pp. 479–484, Sep. 1977.
- [55] L. Biehl. (2001). *An Introduction to MultiSpec*, West Lafayette, IN: School Electr. Comput. Eng., Purdue Univ. [Online]. Available: <http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/>
- [56] V. N. Vapnik, *Statistical Learning Theory*. Hoboken, NJ: Wiley, 1998.
- [57] R. G. Congalton and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Boca Raton, FL: CRC Press, 1999.
- [58] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, pt. 2, no. 6, pp. 1870–1880, Jun. 2007.
- [59] L. Bruzzone, M. Chi, and M. Marconcini, *Semisupervised Support Vector Machines for Classification of Hyperspectral Remote Sensing Images. Hyperspectral Data Exploitation Theory and Applications*. New York: Wiley-Interscience, 2007, pp. 275–311.



**Michele Dalponte** (S'07) received the M.S. degree in telecommunications engineering from the University of Trento, Trento, Italy, in 2006, where he is currently working toward the Ph.D. degree.

He is with the Remote Sensing Laboratory, Department of Information and Communication Technology, University of Trento, and also with the Centro di Ecologia Alpina, Trento. His research interests include the field of remote sensing and pattern recognition, in particular on the analysis of hyperspectral and LIDAR data for forest monitoring.



**Lorenzo Bruzzone** (S'95–M'98–SM'03) received the M.S. degree (*summa cum laude*) in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

From 1998 to 2000, he was a Postdoctoral Researcher with the University of Genoa. In 2000, he joined the University of Trento, Trento, Italy, where he is currently a Full Professor of telecommunications and the Head of the Remote Sensing Laboratory, Department of Information and Communication

Technology. He teaches remote sensing, pattern recognition, and electrical communications. He is an evaluator of project proposals for many different governments (including European Commission) and scientific organizations. He is the author or coauthor of 60 scientific publications in referred international journals, more than 120 papers in conference proceedings, and 7 book chapters. He is a referee for many international journals and has served on the scientific committees of several international conferences. His current research interests include the area of remote-sensing image processing and recognition (analysis of multitemporal data, feature selection, classification, regression and estimation, data fusion, and machine learning). He conducts and supervises research on these topics within the frameworks of several national and international projects.

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). He was a recipient of Recognition of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Reviewers in 1999. He was a Guest Editor of a special issue of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING on the subject of the analysis of multitemporal remote-sensing images (November 2003). He was the General Chair and a Cochair of the First and Second IEEE International Workshop on the Analysis of Multitemporal Remote-Sensing Images (MultiTemp), and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. From 2004 to 2006, he was an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He is currently an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is a member of the Scientific Committee of the India–Italy Center for Advanced Research. He is also a member of the International Association for Pattern Recognition and the Italian Association for Remote Sensing.



**Damiano Gianelle** received the M.S. degree in forest sciences and the Ph.D. degree in environmental agronomy from the University of Padua, Padua, Italy, in 1993 and 1997, respectively.

In 1997, he joined the Centro di Ecologia Alpina, Trento, Italy, where he is currently a Scientific Coordinator. He taught some university courses for M.Sc. (Veterinary Faculty, Padova) and some postgraduate courses (Veterinary Faculty, Milano) for two years. He supervised many graduate thesis and Ph.D. works on remote sensing and modeling. He is currently

the coordinator for the institution of the UE projects Carboeurope-IP and Nitroeurope-IP and of the national project Carboitaly. He is the author or coauthor of 10 scientific publications in referred international journals, more than 20 papers in conference proceedings, and 2 book chapters. He is a referee for the *International Journal of Remote Sensing*. His research interests include remote sensing, eddy covariance, carbon balance, ecophysiology, and forest resources inventory. He conducts and supervises research on these topics within the frameworks of several national and international projects.