IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 11, 2018, accepted November 2, 2018, date of publication November 9, 2018,
date of current version December 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2880083

Deep Learning for Fusion of APEX Hyperspectral and Full-Waveform LiDAR Remote Sensing Data for Tree Species Mapping

WENZHI LIAO[1], (Senior Member, IEEE), FRIEKE VAN COILLIE[2], LIANRU GAO[3], (Senior Member, IEEE), LIWEI LI[3], BING ZHANG[3,4], (Senior Member, IEEE), AND JOCELYN CHANUSSOT[5], (Fellow, IEEE)

[1]Department of Telecommunications and Information Processing, IMEC–Ghent University, 9000 Ghent, Belgium
[2]Department of Forest and Water Management, Ghent University, 9000 Ghent, Belgium
[3]Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
[4]University of Chinese Academy of Sciences, Beijing 100094, China
[5]University of Grenoble Alpes, CNRS, Grenoble Images Speech Signals and Automatics Laboratory, F-38000 Grenoble, France

Corresponding authors: Wenzhi Liao (wenzhi.Liao@ugent.be) and Bing Zhang (zb@radi.ac.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 91638201 and in part by the project of the Research Foundation–Flanders (FWO): Data fusion for image analysis in remote sensing under Grant G037115N.

ABSTRACT Deep learning has been widely used to fuse multi-sensor data for classification. However, current deep learning architecture for multi-sensor data fusion might not always perform better than single data source, especially for the fusion of hyperspectral and light detection and ranging (LiDAR) remote sensing data for tree species mapping in complex, closed forest canopies. In this paper, we propose a new deep fusion framework to integrate the complementary information from hyperspectral and LiDAR data for tree species mapping. We also investigate the fusion of either ''single-band" or multi-band (i.e., full-waveform) LiDAR with hyperspectral data for tree species mapping. Additionally, we provide a solution to estimate the crown size of tree species by the fusion of multi-sensor data. Experimental results on fusing real APEX hyperspectral and LiDAR data demonstrate the effectiveness of the proposed deep fusion framework. Compared to using only single data source or current deep fusion architecture, our proposed method yields improvements in overall and average classification accuracies ranging from 82.21% to 87.10% and 76.71% to 83.45%, respectively.

INDEX TERMS Deep learning, remote sensing, data fusion, hyperspectral, LiDAR.

I. INTRODUCTION

In the framework of sustainable forest management, there is a need for reliable data on forest parameters such as tree species composition, stand diversity, forest vitality and timber volume. Currently, data acquisition in many regions, e.g., Flanders (Belgium), is done by time-consuming and labor-intensive field campaigns. The development of automated tree species classification algorithms is a typical example that is not only the area of interest to researchers but likewise to forest organizations and management agencies. Recent advances in the remote sensing technology offer the potential to facilitate and improve this information acquisition. In particular, hyperspectral (HS) images provide a detailed description of the spectral signatures of ground covers, whereas Light Detection And Ranging (LiDAR) data gives detailed information about the height of the same surveyed area. HS imagery, covering the visible, near-infrared and shortwave-infrared bands with wavelength ranging from 0.4μm to 2.5μm [1], can be used for detailed quantitative analyses, e.g., determination of chlorophyll or water content in leaves, thus to discriminate tree species [2]–[4]. LiDAR data, offering the three-dimensional position of each reflecting point, can be applied straightforwardly to estimate tree height or parameters like biomass [2]. Much more information about three-dimensional objects can be provided by full-waveform LiDAR data, as it records the time-varying signal

of laser pulses, hence enabling better modeling of the vertical structure of vegetation stands compared to discrete return LiDAR data [6], [7].

Many studies have investigated the use of either HS image alone [3], [4] or LiDAR data alone [5], [6], [13], [26] for forest applications. The use of single data source might not be sufficient for a reliable decision making, for example, the optical hyperspectral data is unable to provide 3D information of plant height and canopy structures, while LiDAR data can. On the other hand, it is very difficult to differentiate tree species with the same height by using LiDAR data alone, while HS data can. The complementary information from LiDAR data, once combined with spectral information, can provide a more comprehensive interpretation for tree species mapping. Recent advances in sensor techniques benefit the acquisitions of HS and LiDAR data from the same study area, boosting technique development of multisensor data fusion for many practical applications [2], [12]. Holmgren *et al.* [14] exploited fusion of multispectral imagery and LiDAR data for tree species classification. Offering much higher spectral resolution, hyperspectral images provide significant improvements on tree species classification over the multispectral images [12], resulting in the extensive use of hyperspectral imagery for tree species classifications.

Koetz *et al.* [11] exploited the use of Support Vector Machines (SVM) to fuse LiDAR and HS bands for fuel composition classification, reporting better classification performances from fusion than from either single sensor alone. [12] combined HS and LiDAR remote sensing data for the classification of complex forest areas in a novel classification system, where multiple classifiers were used to properly integrate multi-sensor information. A random forest model was presented in [23] to automatically fuse HS and LiDAR data for classification of eight common African savanna tree species. Reference [23] found that some attributes (e.g., tree species height) from one data source once integrated with the complementary attributes (like spectral information) from the other data source can significantly improve the classification performances. Reference [21] presented a kernel learning model to cope with fusion of heterogeneous features from both HS and LiDAR data, where the similarity of different feature sources were modeled by Gaussian kernels (with bandwidths varying for different scale features). Yokoya *et al.* [22] proposed a framework to assess the landscape visual quality by comparing the physical features learned from the joint use of HS and LiDAR data with the human-perception-based expertise. The approach of [20] fused spectral features (the first few principal components of the pre-processed hyperspectral data) and sizes& shapes of individual trees (derived from the LiDAR data after individual tree-crown delineation) for Japanese Complex Mixed Forest. Coillie *et al.* [19] transferred both hyperspectral and LiDAR data to PCA domain, then combined PCA features incrementally from both data, and selected the optimal feature set based on the best accuracy for the final classification. Buddenbaum *et al.* [49] integrated HS and full-waveform LiDAR images for tree species and tree-age classification. Their fusion method firstly transformed the three-dimensional matrices containing mean LiDAR intensity values in voxels above ground into multi-band image files (i.e. full-waveform LiDAR images), then normalized HS and full-waveform LiDAR images into the same scale, finally obtained the classification map by concentrating the two normalized data sources together as the input of classifiers.

In a high-resolution remote sensing scene, the footprint of an landscape object typically contains more than one pixel, indicating that a high spatial correlation is expected between neighboring pixels. Many approaches generate additional spatial (contextual and structural) information from the HS and LiDAR data, and then incorporate them to improve the fusion process. Geometric information (derived from LiDAR data through image segmentation) was integrated with the spectral information (from HS data) for the classification of urban areas in [10]. A simple and directional method for multi-sensor data fusion is to concentrate multiple feature sources together as the input of a classifier, which are widely used in remote sensing. For example, in [24] and [28], contextual and structural information were first extracted from both HS and LiDAR data by using morphological attribute filters (MAFs) [15], then fused in a stacked architecture for classification. Despite its convenience, fusion by simply stacking multiple feature sources together may lead to the curse of dimensionality problem and excessive computation time [29]. A graph-based fusion method [29] and its generalized version [30] were proposed to couple dimensionality reduction and feature fusion of the spectral information (of the original HS image) and morphological features (built on both HS and LiDAR data). Reference [31] proposed a new framework to fuse HS and LiDAR data for classification of cloud-shadow remote sensing scenes, where spectral and geometric features (modeled by MAFs) were joint used together to perform classification separately in the cloud-free (classifier is trained by the available training samples) and cloud-cover regions. To obtain the new samples to train classifier under cloud regions, they assumed that multiple features sharing the same class clusters. Reference [33] proposed a total variation based fusion method to project high-dimensional multiple features (extracted by extinction filters [34]) into a lower feature space, while preserving smoothness and spatial structures. Multiple vegetation indices were derived from both HS and LiDAR data, and jointly used together as the input of the random forests classifier in [35] for dryland shrub characterization. Object-oriented methods [18] were developed to fuse HS and LiDAR data for the segmentation of tropical forest remote sensing scenes. Per-pixel (e.g., vegetation index) and object-based features were extracted and integrated in a parallel way for HS LiDAR data fusion [29], in which unsupervised and supervised classification approaches were combined.

However, fusing multi-sensor data for tree species mapping remains challenging. To achieve better performances, conventional fusion methods typically extract and select

discriminative features from multi-sensor data by exploiting signal processing technologies and data analysis strategies, such as principal component analysis [29], morphological operators [30], image segmentation [10], spectral metrics (water absorption, various vegetation indices, etc.), as well as domain specific features of the sensor data (e.g., gradients and flatness profile for LiDAR). Extracting these features often requires domain expertise, prior knowledge, and human labor, e.g., fine-tuning of parameters for image segmentation or the availability of specific spectral bands for spectral metrics (e.g., the vegetation indices). Furthermore, optimizing the combination of multiple handcrafted feature sources (features that are extracted from separate images according to a certain manually predefined algorithm based on the expert knowledge, e.g., vegetation indices, morphological features, etc.) [17] and feature extraction methods increases computational complexities. Deep neural networks [39], [41], also known as deep learning, have recently attracted increasing attention in many applications [39], [41], [50]–[52], ranging from image/video processing and classification, to text/audio classification. Deep learning, as its powerful learning ability, can automatically discover an intricate structure of the data. Multiple level features can be learned from the raw data by deep learning layer by layer, of which the whole process includes feature extraction, feature selection, feature fusion and classification. Therefore, deep learning may provide an alternative solution to not only learn joint feature representations from raw data sources automatically, but also fuse decisions adaptively through its deep-layered structure for the final decision making. Its recent applications in remote sensing can be found [25], [36], [38], [40].

The main objective of this paper is to analyze deep learning fusion of multi-sensor data (HS and single band LiDAR, HS and full-waveform LiDAR data) for tree species mapping. Different from image level data fusion (e.g., pansharpening [53]: fusion of pan/multispectral and hyperspectral data to enhance the spatial resolution of hyperspectral data), this paper focuses on feature level data fusion (aiming at fusing multi-sensor data for better classifying different tree species). Specifically, we analyze current deep feature fusion architecture (by stacking multi-sensor data together as the inputs of deep neural network) for tree species mapping. Despite of the simplicity of such deep feature fusion methods, we show in this paper that current deep feature fusion architecture leads to some unexpected and undesirable results for multi-sensor data fusion, and it is often better not to use it. Therefore, we propose a two-stage deep learning method for fusion of HS and LiDAR data. Different from current deep fusion methods, our proposed method first obtains the possibility maps from each data source (first stage). In the second stage, we then learn the joint representation of multi-sensor data by feeding both obtained possibility maps as the input to the deep neural network. We also investigate the classification performances that can be achieved by fusion of single-band LiDAR and hyperspectral data, and by fusion of full-waveform LiDAR and HS data. Last but not least, we exploit the use of the multiple scale features to see the performances on multi-sensor data fusion, as well as to analyze the diameters of tree species crown. Experimental results show that using a single data source, hyperspectral data produces the most accurate species maps, while full-waveform LiDAR outperforms single-band LiDAR data. With the proposed two-stage deep learning fusion method, we demonstrate that fusion of single-band LiDAR data with HS image has significant improvement over current fusion methods even on fusion of full-waveform LiDAR and HS image. The rest of the paper is organized as follows. In Section II, we introduce the study area and remote sensing data. Section III details the fusion method, including current deep fusion framework and the proposed fusion method. Experimental results and analysis are discussed in Section IV. Conclusions are drawn in Section V along with the future work.

II. STUDY AREA AND REMOTE SENSING DATA

A. STUDY AREA

The study area with central point coordinates 51°4′3.51″N – 3°2′21.35″E is located in the forest reserve Wijnendale in the western part of Belgium, see Figure 1. This forest reserve belongs to a 280 *ha* large forest area and covers approximately 66 *ha*. The forest reserve exists since 1996 and is subject to an intensive monitoring programme of the Institute of Nature and Forest Management (INBO). The agency's current policy features removal of exotic species and promoting spontaneous development processes, which should result in close-to-nature forest. Species composition (see Table 1) is dominated by common beech (*Fagus sylvatica* L.), copper beech (*Fagus sylvatica* L., 'Atropunicea'), pedunculate oak (*Quercus robur* L.), common ash (*Fraxinus excelsior* L.), European larch (*Larix deciduas* Mill.), poplar (*Populus spp.*), and sweet chestnut (*Castanea sativa* Mill.). Sycamore maple (*Acer pseudoplatanus* L.) and alder (*Alnus spp.*) occur in the understorey. The forest is characterized by a high crown closure, non-existence of a pre-ordered spatial tree distribution, growth stage diversity and multi-layering of the canopy.

B. FIELD REFERENCE DATA

In 2003, field data were collected in the framework of the monitoring programme in 121 sample plots that are located on alternate grid points in a systematic grid of 50 *m* by 50 *m*. A follow-up visit was conducted in 2013 to check for any changes. One plot covers an area of 0.1 *ha*. Within each sample plot, tree species, diameter at breast height (DBH) and tree coordinates were recorded for all trees with DBH $\geq$ 5 *cm* [8]. In addition, all trees in the forest reserve with a DBH above 18 *cm* were measured in order to capture all large trees in the forest reserve. In total, 1543 trees were recorded, of which a total of 1450 trees were labeled for the seven species. Tree distribution in the uppercanopy was 27.6% common beech, 5.5% copper beech, 20.6% pedunculate oak, 4.6% common ash, 8.2% european larch, 28.6% poplar and

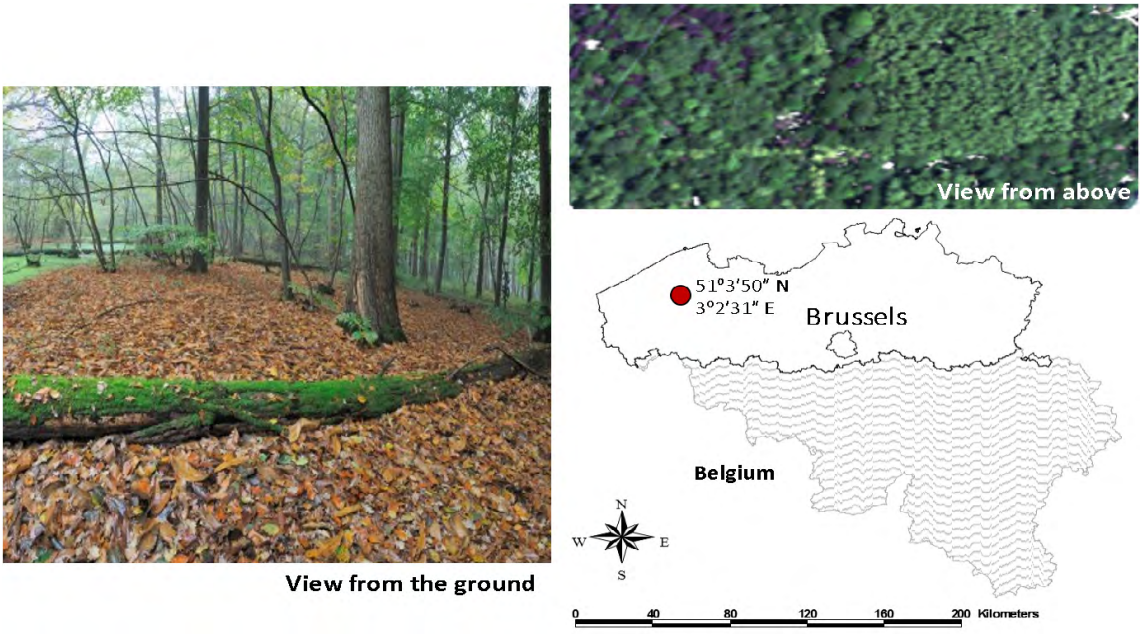


FIGURE 1. Geographical location of Wijnendale forest.

TABLE 1. Training and test samples for remote sensing data.

Class number	Tree species	Mapping color	# Training set	# Test set
1	Beech		80	321
2	Ash		13	54
3	Larch		23	93
4	Poplar		83	333
5	Copper beech		16	64
6	Chestnut		13	54
7	Oak		60	243
	Total		288	1161

4.6% sweet chestnut. Around 20% samples were used for training, the remaining samples were used for testing.

C. REMOTE SENSING DATA

Hyperspectral data was acquired in cloud-free conditions on the 21th of June 2010 with the Airborne Prism Experiment (APEX), with the spatial resolution of 1.5 m. The APEX sensor is the developed imaging spectrometer by a Swiss-Belgian consortium on behalf of the European Space Agency (ESA).[1] The spectral range of APEX sensor is quite broad, covering the visible and short wave infrared wavelength ($372 - 2498$ nm). The total spectral bands are over 300, of which 286 spectral bands were used after removing some noisy bands in the blue part of the electromagnetic spectrum,

[1] www.apex-esa.org

Figure 2(a) shows three band compositions of original hyperspectral data. The radiative transfer model MODTRAN4 [9] was applied to atmospherically correct the radiance values to the top of canopy reflectance. Geometric correction was based on direct georeferencing [16]. Due to sensor problems, no data was registered in two lines of five pixels.

The LiDAR data (TopoSys sensor Harrier 56) was obtained 45 days later on the 4th of August 2010 with a point density of 13.81 m^2 and point spacing of 0.27 m (using all returns). A DTM and DSM were derived at a spatial grid of 0.5 $m \times 0.5$ m. To obtain DSM, we select the maximum of all pulse returns in each grid, then apply a morphological closing filter (with structure element of circular, on size 3×3 pixels) on them. We got DTM by using progressive morphological filter (as proposed in [27]) to remove non-ground features (e.g. trees) from the LiDAR point cloud. The

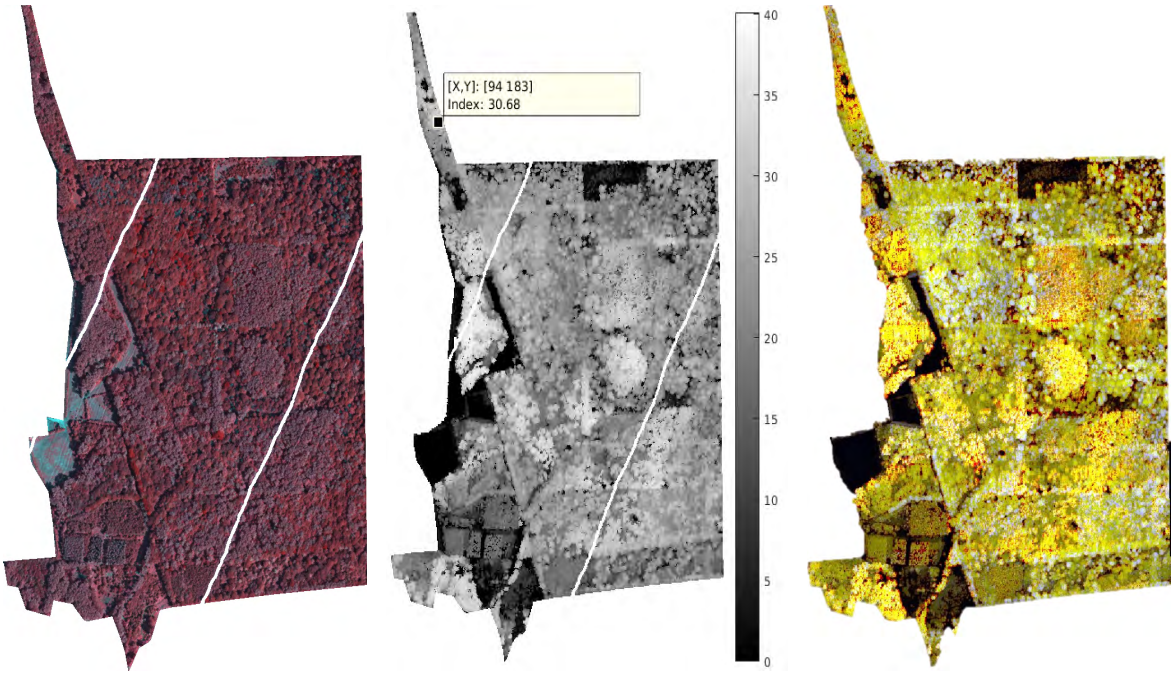


FIGURE 2. Remote sensing data (a) False color image of APEX hyperspectral image; (b) LiDAR data: vegetation height model (the intensity denotes the height of the tree); (c) Percentile height RGB image of full-waveform LiDAR data at levels 100%, 50%, and 0%.

LiDAR data were processed in LAS format and converted to a raster grid with a cell size equal to the spatial resolution of the hyperspectral sensor (1.5 *m*). Two products were obtained: a vegetation height model (VHM) and percentile height (PH) values, see Figure 2(b) and Figure 2(c). The latter was represented as a multi-band image of 11 PH bands (i.e. full-waveform LiDAR data). Height percentiles were computed from the terrain corrected point cloud and represent the height at $x\%$ of all vegetation returns. Low height percentiles (e.g. 10% or 20%) can be good indicators of crown base height. The crown base height might be a good indicator to differentiate tree species. The spectra and height profile of some classes are similar (see Figure 3(a)- 3(b)), challenging separation of tree species by single data source. Tree heights range from 5 to 39 meters, with a median height of 26 meters, see Figure 3(c).

III. METHODOLOGY

The simplest way is to concatenate the different data sources together as an input for classification (see Fig. 4(b)), similar as [28]. However, there are some limitations, such as the curse of dimensionality, excessive computation time, as analyzed in [29] and [30]. Applying feature reduction techniques on all the feature sources together does not take into account the differences in structure of the feature spaces, while dimension reduction on each feature source may lead to hard decision on optimizing the dimension combination [30].

HS and LiDAR data contain complementary information for the same study area, they have correlations at different feature levels. Recently, deep learning [39] has been exploited for joint feature representations. In general, the goal of deep learning is to learn multiple levels of features or representations of the data, where higher level features are derived from lower level features to form a hierarchical representation. This can be achieved through variations of the deep learning algorithms, including deep belief networks (DBNs) [41], stacked autoencoders (SAEs) [39], and stacked denoising autoencoders (SDAEs) [43], deep Boltzmann machines (DBMs) [42], convolutional neural networks [44], and deep sparse filters (DSFs) [45]. In this paper, we focus on stacked autoencoder [39] to learn high-level features of HS and LiDAR, as well as their joint presentation. Autoencoders are unsupervised neural networks (as well as unsupervised feature learning algorithms) that use machine learning for the purpose of dimensionality reduction. Autoencoders can be stacked to form a deep autoencoder network. Due to its relatively simple structure, SAE has been effectively adapted to remote sensing image processing and reported with powerful feature representation capabilities. Chen *et al.* [40] combined the spatial features and spectral information of hyperspectral images with SAE, and achieved a competitive performance for the classification task. In [46] and [47], SAE was adapted to fuse multiple features for SAR image classification. These methods first generate additional texture features on the original SAR image, then flatten them into one-dimensional vectors and feed them into SAE after preprocessing.

Suppose we have training set $T = \{(x_1, y_1), \cdots, (x_n, y_n)\}$, where x_i is the ith sample with its label y_i. An autoencoder (as shown in Figure 5(a)) is a type of neural network that is widely used for feature learning and dimensionality

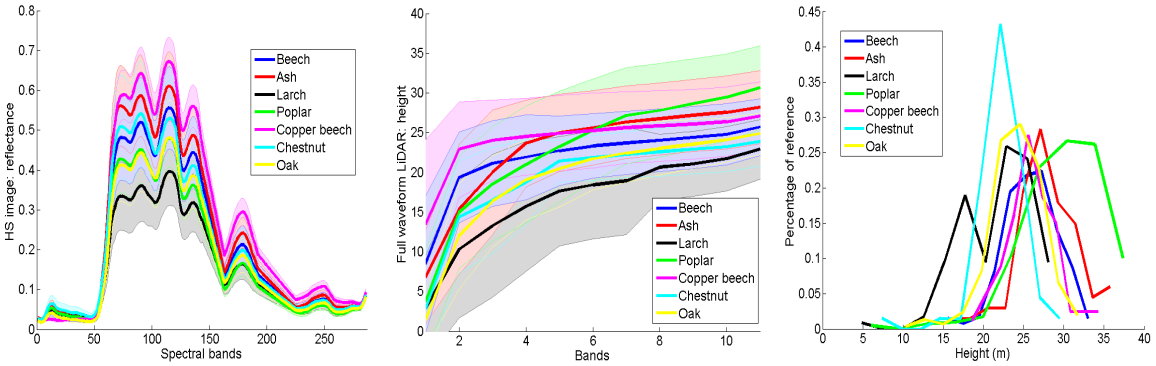


FIGURE 3. Remote sensing data. (a) Mean spectra (solid line), and standard deviation (error bar) with shaded area; (b) Mean height above the ground (solid line), and standard deviation (error bar) with shaded area; (c) Histogram of reference tree heights for different species.

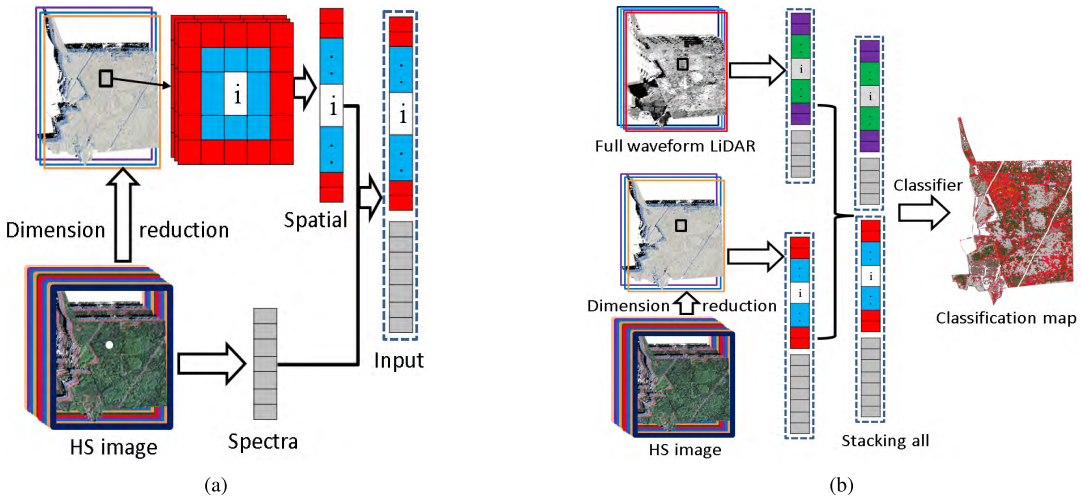


FIGURE 4. Joint feature representations. (a) Spectral-spatial features. (b) Feature concentration.

reduction [48]. It takes an input $x \in \mathbb{R}^D$ and first encodes it to the hidden layer and learns the latent representation $h \in \mathbb{R}^d$ using an activation function $h = f(Wx + b)$ with parameters $\theta = \{W, b\}$, where $f(\cdot)$ is a mapping function such as sigmod function or tanh function. The latent representation is then decoded to an output layer, where the input can be reconstructed by a reverse mapping of $f : \widehat{x} = f(W'x+b')$ with $\theta' = \{W', b'\}$. The two parameter sets for encoding the input and decoding the latent representation are usually constrained with $W' = W^T$, i.e., the same weights for both mappings of input-to-hidden layer and hidden-to-output layer. For each training pattern $x_i(\forall i \in [1, n])$, it can be mapped onto its hidden code h_i and its reconstruction $\widehat{x}_i$. The parameters can be optimized, by minimizing the reconstruction error between input x and reconstruction $\widehat{x}$ through an appropriate cost function (e.g., mean squared error or cross entropy error) over the training set. A stacked autoencoder [41] has multiple hidden layers of neurons between the input and output layers, see Figure 5(b). A good solution to obtain the optimal parameters for a SAE is the use of greedy layer-wise training. To obtain the parameters of the first layer $\theta_1 = \{W_1^k, b_1^k\}$, $k \in \{1, 2\}$, we take the inputs and feed them into SAE for training, transforming the raw inputs into a vector consisting of activation of the hidden units. The vector is then used to train the second layer and obtain the parameters $\theta_2 = \{W_2^k, b_2^k\}$, $k \in \{1, 2\}$. The output of each layer is used as the input for the subsequent layer to complete the pre-training. Fine-tuning with back propagation is typically explored to ensure the network converges to a global minimum, improving the final performances.

A. JOINT FEATURE REPRESENTATION BY DEEP LEARNING

The high dimensionality of the remote sensing data (especially for hyperspectral data) makes the generation of neighborhood information (with a sliding window centered per pixel) exhaustive for computer resources. Besides, high dimensionality also involves redundancy between the HS bands. To overcome these problems, feature extraction (we use PCA: Principal Component Analysis in this paper) is firstly used to reduce the redundancy as well as the

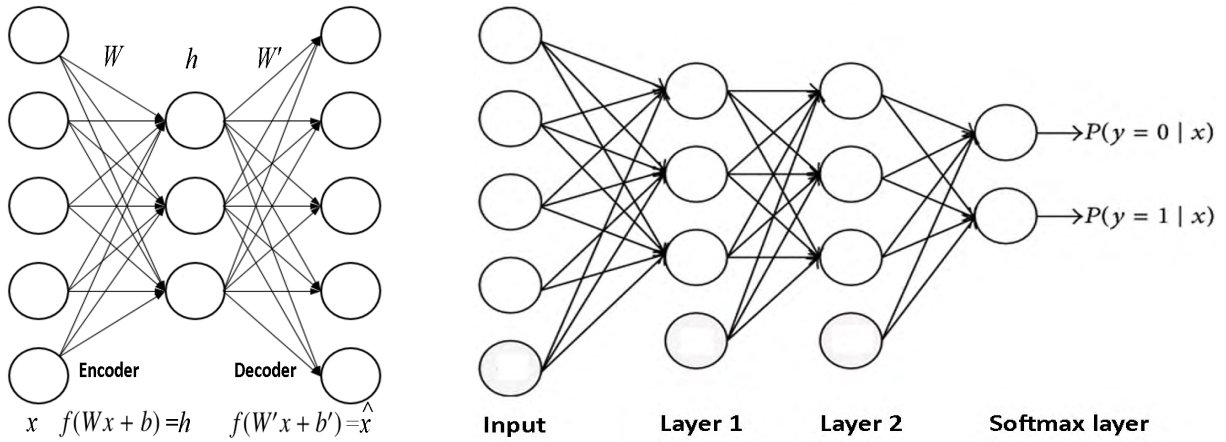


FIGURE 5. Structure of autoencoder and stacked autoencoder. (a) A three-layers autoencoder; (b) A stacked autoencoder composed of two hidden layers (i.e. autoencoders).

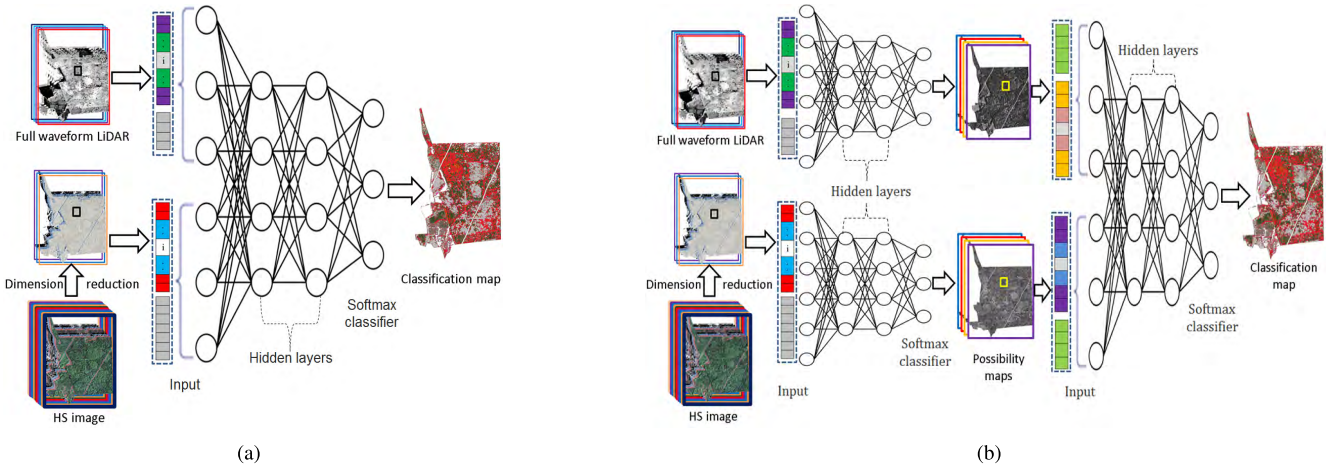


FIGURE 6. Joint feature representation by deep learning. The input is the flattened vector of a region centered at a pixel, same as the spatial in 4(a). (a) Current deep feature fusion framework. (b) Proposed two-stage deep fusion framework.

dimensionality of the hyperspectral data. In particular, the first 12 principal components were selected (representing 99.8% of the cumulative variance), and then the neighborhood information (a centered pixel with its neighboring pixels within a sliding window) is generated on these few extracted features, as shown in Figure 4(a), similar as the approaches of [36], [40], [46], and [47].

In order to learn joint feature representations from multi-sensor data by deep learning, current deep fusion methods typically use the raw data and their additional information (generated on the raw data) as the pre-trained layers [36], [40], [46], [47] (Fig. 6(a)). In particular, the outputs of the first layer vectors are used as the input to train the subsequent new layer. The representations from the first layer correspond to local pixel regions, whereas the second or higher layer features (also called deeper features) model the relationships between them. By learning joint representions through multiple layers, it becomes easier for the model to learn invariant information and higher-order correlations between multi-sensor data.

Due to the simplicity of such deep feature fusion model (simply concatenate several kinds of raw features together as the inputs of deep neural network), it is widely used in remote sensing [36] and [40]. However, the performances of this deep fusion model may not be better (or even worse) than using single data source. This is because the element values of different data sources can be significantly unbalanced (dimension unbalance), especially for fusion of HS image and single band DSM, the dimensionality of features from HS image is more than 500, whereas the dimension of features from DSM is just 26 (if we use a sliding window with size of 5×5). The information contained by different data source is not equally represented. When calculating the distance between two vectors (either for mean square errors or similarity), the much lower-dimensional features will make less contribution for the decision of the vector difference, even

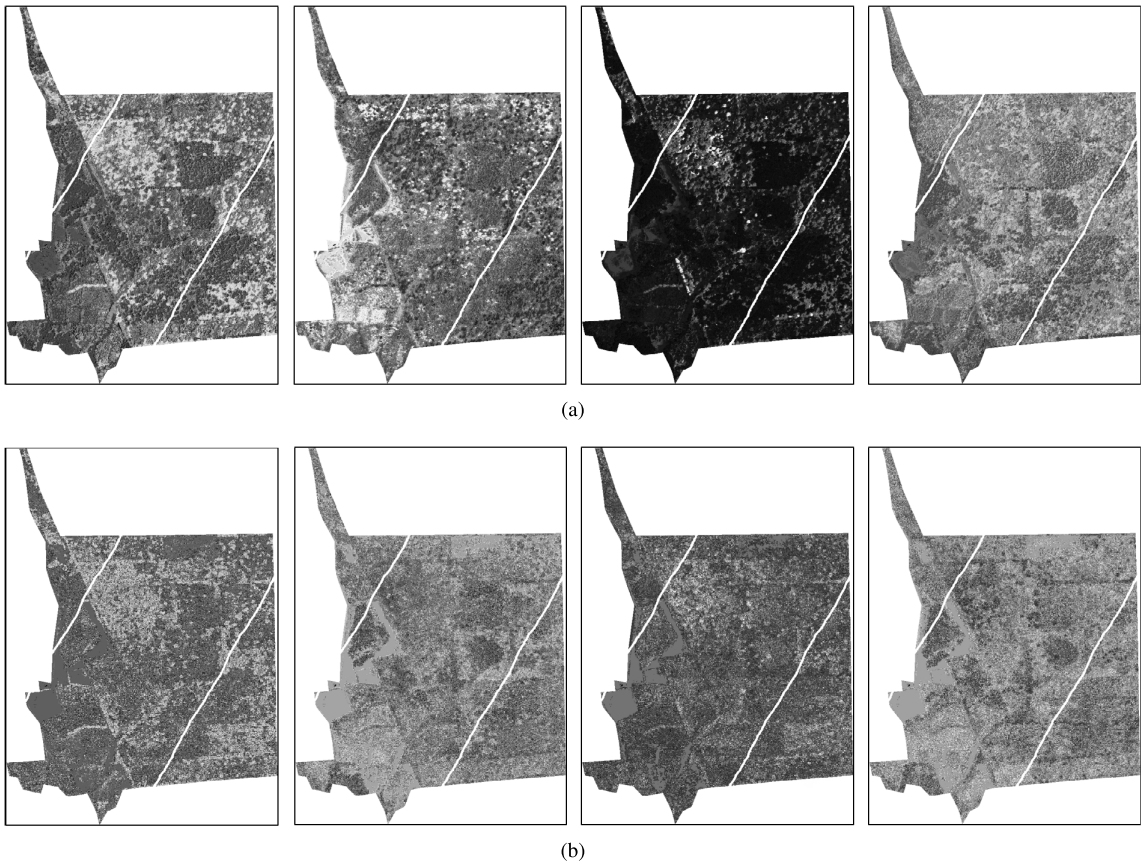


FIGURE 7. Possibility maps of remote sensing images by deep learning. (a) Possibility maps of HS image, possibility channel 1, 3, 5 and 7. (b) Possibility maps of LiDAR image, possibility channel 1, 3, 5 and 7.

though they are more discriminative. Furthermore, Multi-sensor data are usually with different ranges and standard deviations (see Figure 3(a) and 3(a)) and that might bias the training. Last but not the least, the inputs of SAE by stacking raw data and their neighbors contain redundant information. Current deep fusion models, either in image processing [51], [52] or computer vision [50], assume multi-sensor data have similar dimensions. For fusion of multi-sensor data with largely unbalanced features, some deep feature fusion methods produce similar performances as using the raw data, as discussed in [54].

To overcome the limitations, we propose a two-stage deep fusion framework to learn joint features of HS and LiDAR data for tree species mapping, Figure 6(b) shows the proposed flowchart. In particular for the first stage, the proposed method uses each raw data source and its neighbors (within a sliding window) as the pre-trained layer (the spectral-spatial features in Figure 4(a)) of SAE to obtain individual possibility maps for each data source, the fusion architecture is similar as [40]. In the second stage, both obtained possibility maps and their neighbors are concatenated as the inputs of a SAE to learn the joint feature representation, similar to the method of [36], but using stacked auto-encoder instead of deep sparse filters.

This way the element values of different data sources are much better balanced. The dimensions of both possibility maps are equal to the number of the tree species (here we have 7 tree species in the remote sensing scene), Figure 7(a)-7(b) displays some possibility maps obtained by deep learning on each data source, respectively. Moreover, the intensity scales of both possibility maps are the same, ranging from 0 to 1. Additionally, the resulting possibility maps obtained by different data sources have the same statistical meaning, i.e., larger value (brighter in the possibility maps) of a pixel in the channel of possibility maps means higher possibility belonging to the specific class (i.e. the index of possibility maps).

From the possibility maps, we can see that the distributions of bright pixels are different from the results of two different data sources. For example in Figure 7(a)-7(b), those bright points in the channel 5 of HS possibility maps are much higher possibility to class 5; whereas in LiDAR possibility maps, these points may belong to other classes (since their intensities are relatively smaller). From the average possibility profile and their error bars in Figure 8(a)-8(b), classes of 'Ash', 'Larch' and 'Poplar' are much easier to discriminate by using the HS image than LiDAR data, as their possibilities are much higher in their corresponding indexes (around 50%).

TABLE 2. Classification accuracy (%) using single data source.

	SVM						Deep Learning					
Class	HS	DSM	FuLi	S-HS	S-DSM	S-FuLi	HS	DSM	FuLi	S-HS	S-DSM	S-FuLi
1	80.55	79.05	83.04	83.79	78.05	**85.79**	73.82	53.87	68.58	79.55	53.37	76.06
2	52.24	0.0	22.39	61.19	4.48	16.42	**68.66**	4.48	29.85	**68.66**	26.87	23.88
3	83.62	0.0	26.72	80.17	21.55	42.24	**87.93**	25.0	36.21	84.48	42.24	37.07
4	93.99	64.42	66.35	**96.15**	67.79	85.34	93.51	60.58	79.33	92.31	66.35	82.69
5	**100.0**	0.0	3.57	85.00	5.0	20.0	90.0	6.25	31.25	93.75	28.75	37.50
6	47.76	0.0	2.99	37.31	1.49	13.43	**53.73**	4.48	20.9	35.82	23.88	29.85
7	69.97	0.0	70.63	75.58	24.75	74.59	69.31	36.3	65.35	**76.57**	43.56	61.39
OA	80.69	40.34	60.28	**82.21**	48.48	69.66	79.38	42.62	62.34	81.24	50.21	65.10
AA	75.45	20.50	39.41	74.17	29.02	48.26	**76.71**	27.28	47.35	75.88	40.72	19.78
κ	75.12	17.2	47.61	**76.89**	30.02	59.81	73.65	23.59	51.28	75.97	36.38	54.48
Time (s)	162.31	**7.67**	19.13	167.98	26.86	160.24	584.36	89.05	153.51	532.18	261.61	489.85

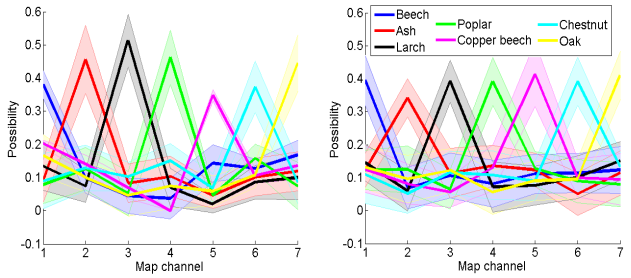


FIGURE 8. Possibility maps: Mean possibility profile (solid line) and standard deviation (error bar) with shaded area of (a) HS image; (b) full-waveform LiDAR image. Note: Map channel means the dimension of the possivility maps.

As the inputs of deep neural network, the possibility maps are more discriminative than by using the raw data, see the raw spectra and height profile from the raw full-waveform LiDAR data in Figure 3(a)-3(b). Furthermore, the resulting inputs contain much less redundant information than raw data.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To evaluate the proposed fusion method for multi-sensor remote sensing data classification, we compare the following schemes:

1) Using the raw data source, e.g., raw APEX hyperspectral image (HS), raw single band LiDAR image (DSM), raw full-waveform LiDAR image (FuLi).
2) Combining raw data and its neighborhood information, e.g., APEX HS data (S-HS), single band LiDAR image (S-DSM), full-waveform LiDAR image (S-FuLi).
3) Stacking several data sources and using them as input of support vector machine classifiers (SVM), similar fusion framework as [28].
4) Transferring both hyperspectral and LiDAR data to PCA domain, PCA features were incrementally combined and the optimal feature set was selected based on the best OA [19], named PCA_{Fu}.
5) Deep feature fusion methods (Deep SAE), similarly as [36] and [40].
6) A recent method proposed for hyperspectral tree species classification by combining LiDAR data [20], named Mat15. Note that: to make the fair comparison for all methods, we did not apply post-processing on the final classification maps.

SVM [55] classifier performs well even with a limited number of training samples. The SVM classifier with radial basis function (RBF) kernels in Matlab SVM Toolbox, LIBSVM [56], is applied in our experiments. We apply a grid-search on optimizing the parameters of SVM C (the penalty factor) and γ (the RBF kernel widths) by using 5-fold cross-validation to find the best C within the given set $\{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ and the best γ within the given set $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$. In this paper, the parameter settings of stacked autoencoder are: learning rate 0.1 and training epochs 2000, 3 layers and 120 hidden units per layer. The same training samples were used to train both stages of our proposed method. To quantitatively evaluate the classification results, we calculate the Overall Accuracy (OA), the Average Accuracy (AA) and the Kappa coefficient (κ) over the test samples. OA can be calculated by the number of well-classified samples divided by the number of all test samples, AA is sum of accuracy of each class divided by the number of classes. Kappa can be calculated from the observed and expected frequencies on the diagonal of a square contingency table [32]. It measures the percentage of data values in the main diagonal of the table and then adjusts these values for the amount of agreement that could be expected due to chance alone. Table 2 and Table 3 report the results by using single data sources and fusion of both data, respectively. We compared the computational cost of different approaches. All the methods were implemented in Matlab 2015b. The experiments were carried out on 64-bit,

TABLE 3. Classification accuracy (%) by fusing HS and LiDAR data.

Class	Fusion of DSM and HS data				Fusion of full-waveform LiDAR and HS data				
	SVM	Deep SAE	Mat15 [20]	Proposed	SVM	PCA_{Fu} [19]	Deep SAE	Mat15 [20]	Proposed
1	79.30	68.33	80.30	83.04	**88.53**	87.28	77.31	84.04	84.54
2	**76.12**	55.22	62.69	74.63	70.15	64.18	55.22	61.19	71.64
3	85.34	72.41	87.07	**87.07**	86.21	86.21	53.45	82.76	**87.07**
4	**96.63**	88.22	94.47	95.67	**96.63**	93.03	90.38	96.39	93.27
5	93.75	87.50	92.50	86.25	85.00	88.75	87.50	85.00	**95.00**
6	38.81	47.76	58.21	49.25	32.84	53.73	50.75	50.75	**64.18**
7	73.60	72.94	72.28	79.21	78.88	80.20	72.31	82.84	**88.45**
OA	82.34	74.83	82.07	84.41	85.03	84.83	76.48	84.69	**87.10**
AA	77.65	70.34	78.21	79.30	76.89	79.05	69.60	77.58	**83.45**
κ	77.22	67.84	76.92	79.93	80.59	80.41	69.77	80.21	**83.51**
Time (s)	**313.01**	605.52	1589.36	1102.03	353.62	7826.25	716.69	2356.71	1341.09

3.20GHz Intel i7-3930K (4 core) CPU computer with 32-GB memory. Table 2 and Table 3 report the computational time of different approaches. The recorded times were consumed in both training and test stages, including the time consumed on the parameter determination of some methods (such as C and γ optimization in SVM, parameters tuning in stacked autoencoder). We can see that fusion methods based on SVM schemes are faster than deep learning based fusion methods. The proposed two-stage produces better results but consumes more time than deep SAE, due to the parameters tuning in two stage training of the networks. The consumed time will increase as the dimensionality increases.

A. SINGLE DATA SOURCE CLASSIFICATION

The results show that for single data source, APEX HS data produces better OA, AA and κ than either single band DSM or full-waveform LiDAR data, with 20%-40% improvements in terms of OA. The improvements by using the full-waveform LiDAR data over single band DSM are significant, with OA higher than almost 20 percentage points. A digital surface model (DSM) provides an estimate of the tree canopy height. Full-waveform LiDAR data have the potential to provide much richer spatial information about canopy characteristics in three dimensions than single band DSM. This is because full-waveform airborne laser scanning can model the vertical structure of vegetation stands [6], by measuring the time-varying signal of the laser pulse. By integrating the neighboring information (through a sliding window centered per pixel), the improvements on classification performance of each data source are significant, especially for LiDAR data, with OA improvements of 3%-9% over the raw LiDAR data. The schemes by using SVM classifier produce similar accuracies as deep learning schemes for single data source.

From the class-specific accuracies, APEX hyperspectral data has advantages to discriminate most of classes, especially for tree species 'Poplar' and 'Copper beech', with accuracies higher than 90%. Raw single band DSM misclassifies most classes, producing 0% accuracies for 'Ash', 'Larch', 'Copper beech', 'Chestnut' and 'Oak'. For class 'Beech', full-waveform LiDAR data with SVM classifier produces better accuracy than the other data sources.

B. FUSION OF MULTI-SENSOR DATA FOR CLASSIFICATION

It can be found that single HS or LiDAR data is not sufficient for a reliable tree species mapping (especially for class 6), the complementary information for some specific classes is evident. However, simple fusion of both data sources won't always be better than using single data source. For fusion of single band DSM and HS data, the fusion architecture by simply stacking HS and LiDAR feature sources as input of SVM classifier has not significant improvement over using single HS features. Current deep feature fusion architecture, similarly as the approaches proposed in [36] and [40], produces much worse results than using single data source, dropping by 4-6 percentage points compared to using single HS features. In particular, the performance drops significantly on tree species 'Beech', decreasing from 85.79% (by using S-FuLi) to 68.33% (by fusion of single DSM and HS data) and to 77.31% (by fusion of full-waveform LiDAR and HS data); whereas for class 'Larch', the accuracy drops from 87.93% (by using HS with deep learning) to 72.41% (by fusion of DSM and HS data) and to 55.22% (by fusion of full-waveform LiDAR and HS data).

Fusion of full-waveform LiDAR data and HS data outperforms fusion of DSM and HS data. Compared to the situation with the fusion of DSM and HS data, the OA of fusing full-waveform LiDAR and HS data has 2.69%, 1.65%, 2.62% and 2.69% improvements for SVM scheme, current deep fusion framework, Mat15 [20] and proposed two-stage deep fusion method, respectively. The proposed fusion method takes full advantages of both data sources, while addressing the shortcomings of each of them, and thus enables better performances. With the proposed method, the improvements of OA over SVM scheme, PCA_{Fu} [19], current deep fusion

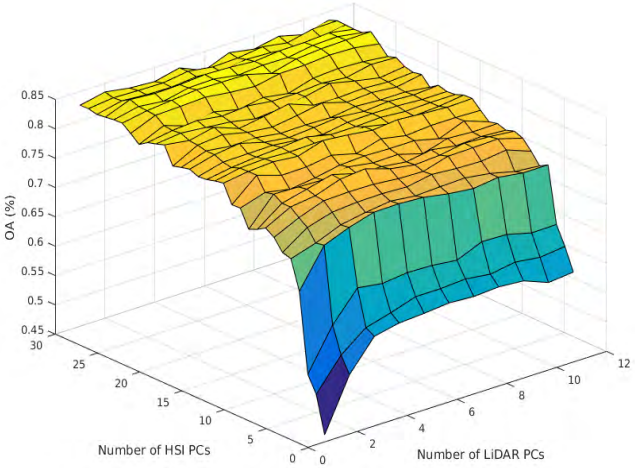


FIGURE 9. Surface of the OA as a function of the number of the principal components (PCs) for both hyperspectral and LiDAR data.

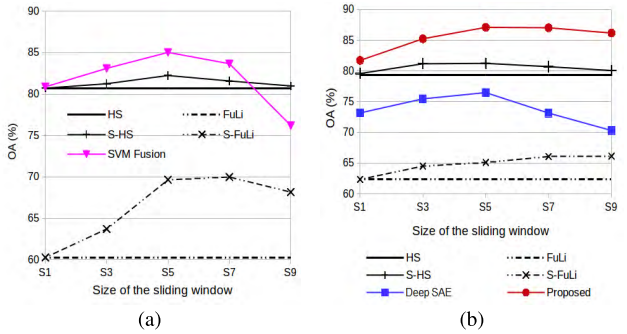


FIGURE 10. Classification accuracy as a function of the sliding window size. S3 denotes the size of sliding window is 3 × 3. (a) SVM classifier. (b) Deep learning.

framework and Mat15 [20] are around 2%, 3%, 10% and 3%, respectively.

From Figure 9, it confirms that combining hyperspectral and LiDAR data will increase the tree species classification accuracy, as was also reported in [19]. The classification accuracy will first increase then keep stable as the number of the principal components of either hyperspectral or LiDAR data increases. It is not easy to select the optimal combination of the feature set from multi-source data. PCA_{Fu} [19] took more than two hours to search the optimal combination of the principal components from hyperspectral and LiDAR data. However, this combination will change as the training samples (or training size) change.

C. FUSION FOR ESTIMATION OF TREE SPECIES CROWN

The crown size of tree species is a key parameter in this context as it correlates with the space a tree species occupies. The changes of the crown size indicates the competition between tree species in a forest. Several methods have been developed to estimate the crown size, with many prior knowledges known in advance [37], [57], [58]. We propose a method to estimate the diameter of tree species crown, by exploiting fusion of multi-scale features from hyperspectral and LiDAR data. Multi-scale features are generated by applying a sliding window centered at one pixel on both HS and LiDAR data. The scale size is denoted by the size of sliding widow. Each tree species has their own diameter, which can be indicated by the scale size. The size of the sliding window has a significant influence on neighboring information modeling. On the one hand, when the window size is too small, the neighborhood region contains too few samples to properly model the local spatial information of the centered pixel, failing to discriminate different tree species. On the other hand, if we set the sliding window size too large, then we might not retrieve the local spatial information (imagine the case when the window size set to the size of the whole image). Figure 10(a)- 10(b) shows the performances of different fusion schemes as the size of sliding window increases.

The results confirm that incorporating spatial information improves the performances of tree species mapping. It is very interesting to find that as the size of sliding window increases, the accuracy first increases then decreases. We can indicate that the diameters of most tree species crown in HS scene are around 9-15 meters (note: the spatial resolution of remote sensing data is 1.5 meters), as it can be seen from Figure 10(a)- 10(b) that the accuracy reaches the top when using scale sizes of 3 × 3 and 5 × 5. The diameters of height layers (for the crown of most tree species) in full waveform LiDAR scene are around 15-21 meters, as we can see obviously the accuracies keep stable when the scale sizes are 5 × 5 and 7 × 7. Therefore, the fusion method produces the best result at the scale 5 × 5, the complementary information of both HS and LiDAR data is well exploited at this scale.

It is better to use single feature source (either S-HS or S-FuLi features) for tree species mapping when the scale size is larger than the radius of most tree species crown. We can see that the results by fusing both feature sources decreases significantly when the scale sizes are larger than 5 × 5, especially for SVM fusion schema. On the other hand, we can estimate the diameter of most tree species crown, through fusion of hyperspectral and LiDAR data.

V. CONCLUSIONS

This paper exploited deep learning to fuse hyperspectral and LiDAR data for tree species mapping. In particular, we developed a two-stage deep learning fusion method for multi-sensor data fusion. The results on fusion of APEX hyperspectral and LiDAR data from complex, closed forest canopies in the western part of Belgium, show that the proposed two-stage deep fusion method is effective in integrating multi-sensor data for classification, with significant improvements over current deep feature fusion architecture [36], [40], [50]. Full-waveform LiDAR data produced much better results than single band LiDAR data. Fusion of HS and full-waveform LiDAR outperforms fusion of HS and single band LiDAR data. Fusion multi-sensor data by simply stacking all

data sources together or using current deep fusion architecture won't always have significant improvements over single data source, especially for single band LiDAR data.

ACKNOWLEDGMENT

We gratefully acknowledge The Research Institute for Forest and Nature for collecting and providing the field reference data in Wijnendale. The authors thank the Remote Sensing Laboratories (University of Zurich, Switzerland) and the Flemish Institute for Technological Research (VITO, Belgium) for respectively pre-processing the LiDAR and hyperspectral data. The Belgian Science Policy Office (BELSPO) is thanked for financing this research in the framework of the HYPERFOREST project (STEREO II). Wenzhi Liao is a postdoctoral fellow of the Research Foundation Flanders (FWO) and acknowledges its support.

REFERENCES

[1] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.

[2] F. E. Fassnacht *et al.*, "Review of studies on tree species classification from remotely sensed data," *Remote Sens. Environ.*, vol. 186, pp. 64–87, Dec. 2016.

[3] A. Ghiyamat and H. Z. M. Shafri, "A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment," *Int. J. Remote Sens.*, vol. 31, no. 7, pp. 1837–1856, Apr. 2010.

[4] M. A. Peña, R. Liao, and A. Brenning, "Using spectrotemporal indices to improve the fruit-tree crop classification accuracy," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 158–169, Jun. 2017.

[5] M. Milenković, W. Wagner, R. Quast, M. Hollaus, C. Ressl, and N. Pfeifer, "Total canopy transmittance estimated from small-footprint, full-waveform airborne LiDAR," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 61–72, Jun. 2017.

[6] F. Morsdorf *et al.*, "Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning," *Remote Sens. Environ.*, vol. 114, no. 7, pp. 1403–1415, Jul. 2010.

[7] A. Hovi, L. Korhonen, J. Vauhkonen, and I. Korpela, "LiDAR waveform features for tree species classification and their sensitivity to tree- and acquisition related parameters," *Remote Sens. Environ.*, vol. 173, pp. 224–237, Feb. 2016.

[8] L. De Keersmaeker, H. Baete, B. Christiaens, M. Esprit, P. Van de Kerckhove, and K. Vandekerkhove, "BosreservaatWijnendalebos: Monitoringrapport—Monitoring van de dendrometrische gegevens en de vegetatie in de steekproefcirkels en de kernvlakte," Instituut voor Bosbouw enWildbeheer, Brussels, Belgium, Tech. Rep., 2005, vol. 10.

[9] A. Berk *et al.*, "MODTRAN4 radiative transfer modeling for atmospheric correction," *Proc. SPIE*, vol. 3756, pp. 348–353, Oct. 1999.

[10] 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, USA, Mar. 2005, pp. 1–6.

[11] B. Koetz *et al.*, "Fusion of imaging spectrometer and LIDAR data over combined radiative transfer models for forest canopy characterization," *Remote Sens. Environ.*, vol. 106, no. 4, pp. 449–459, Feb. 2007.

[12] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[13] J. Heinzel and B. Koch, "Exploring full-waveform LiDAR parameters for tree species classification," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 1, pp. 152–160, Feb. 2011.

[14] J. Holmgren, Å. Persson, and U. Söderman, "Species identification of individual trees by combining high resolution LiDAR data with multispectral images," *Int. J. Remote Sens.*, vol. 29, no. 5, pp. 1537–1552, 2008.

[15] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Dec. 2010.

[16] K. M. Brown, G. M. Foody, and P. M. Atkinson, "Modelling geometric and misregistration error in airborne sensor data to enhance change detection," *Int. J. Remote Sens.*, vol. 28, no. 12, pp. 2857–2879, May 2007.

[17] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification," *Pattern Recognit.*, vol. 71, pp. 158–172, Nov. 2017.

[18] G. Tochon *et al.*, "On the use of binary partition trees for the tree crown segmentation of tropical rainforest hyperspectral images," *Remote Sens. Environ.*, vol. 159, pp. 318–331, Mar. 2015.

[19] F. M. B. Van Coillie *et al.*, "Optimized feature fusion of LiDAR and hyperspectral data for tree species mapping in closed forest canopies," in *Proc. 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Tokyo, Japan, Jun. 2015, pp. 1–4.

[20] T. Matsuki, N. Yokoya, and A. Iwasaki, "Hyperspectral tree species classification of Japanese complex mixed forest with the aid of lidar data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens*, vol. 8, no. 5, pp. 2177–2187, May 2015.

[21] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, "A novel MKL model of integrating LiDAR data and MSI for urban area classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5312–5326, Oct. 2015.

[22] N. Yokoya, S. Nakazawa, T. Matsuki, and A. Iwasaki, "Fusion of hyperspectral and LiDAR data for landscape visual quality assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2419–2425, Jun. 2014.

[23] L. Naidoo, M. A. Cho, R. Mathieu, and G. Asner, "Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment," *ISPRS J. Photogramm. Remote Sens.*, vol. 69, pp. 167–179, Apr. 2012.

[24] M. Khodadadzadeh, J. Li, S. Prasad, and A. Plaza, "Fusion of hyperspectral and LiDAR remote sensing data using multiple feature learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2971–2983, Jun. 2015.

[25] P. Ghamisi, B. Höfle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[26] J.-M. Monnet, J. Chanussot, and F. Berger, "Support vector regression for the estimation of forest stand parameters using airborne laser scanning," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 580–584, May 2011.

[27] K. Zhang, S.-C. Chen, D. Whitman, M.-L. Shyu, J. Yan, and C. Zhang, "A progressive morphological filter for removing nonground measurements from airborne LIDAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 4, pp. 872–882, Apr. 2003.

[28] M. Pedergnana, P. R. Marpu, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 856–865, Nov. 2012.

[29] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.

[30] W. Liao, A. Pizurica, R. Bellens, S. Gautama, and W. Philips, "Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 552–556, Mar. 2015.

[31] R. Luo *et al.*, "Fusion of hyperspectral and LiDAR data for classification of cloud-shadow mixed remote sensed scene," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3768–3781, Aug. 2017.

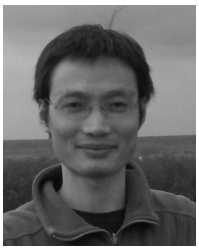
[32] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: The kappa statistic," *Family Med.*, vol. 37, no. 5, pp. 360–363, 2005.

[33] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, Jul. 2017.

[34] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. A. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, Oct. 2016.

[35] J. Mitchell, R. Shrestha, L. P. Spaete, and N. F. Glenn, "Combining airborne hyperspectral and LiDAR data across local sites for upscaling shrubland structural information: Lessons for HyspIRI," *Remote Sens. Environ.*, vol. 167, pp. 98–110, Sep. 2015.

[36] A. Merentitis and C. Debes, "Automatic fusion and classification using random forests and features extracted with deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Milan, Italy, Jul. 2015, pp. 2943–2946.
[37] H. Pretzsch *et al.*, "Crown size and growing space requirement of common tree species in urban centres, parks, and forests," *Urban Forestry Urban Greening*, vol. 14, no. 3, pp. 466–479, 2015.
[38] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
[39] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Neural Inf. Process. Syst.*, Cambridge, MA, USA, 2007, pp. 153–160.
[40] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[41] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[42] R. Salakhutdinov and G. Hinton, "Deep Boltzmann machines," in *Proc. Int. Conf. Artif. Intell. Statist.*, Clearwater Beach, FL, USA, 2009, pp. 448–455.
[43] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010.
[44] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.
[45] J. Ngiam, Z. Chen, S. A. Bhaskar, P. W. Koh, and A. Y. Ng, "Sparse filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, Eds. New York, NY, USA: Curran, 2011, pp. 1125–1133.
[46] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution SAR image classification via deep convolutional autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 12, no. 11, pp. 2351–2355, Nov. 2015.
[47] M. Kang, K. Ji, X. Leng, X. Xing, and H. Zou, "Synthetic aperture radar target recognition with feature fusion based on a stacked autoencoder," *Sensors*, vol. 17, no. 1, p. 192, Jan. 2017.
[48] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
[49] H. Buddenbaum, S. Seeling, and J. Hill, "Fusion of full-waveform lidar and imaging spectroscopy remote sensing data for the characterization of forest stands," *Int. J. Remote Sens.*, vol. 34, no. 13, pp. 4511–4524, Apr. 2013.
[50] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proc. Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011, pp. 689–696.
[51] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014.
[52] L. Jing, L. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, 2017.
[53] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
[54] W. Liao, A. Pižurica, R. Luo, and W. Philips, "A comparison on multiple level features for fusion of hyperspectral and LiDAR data," in *Proc. Int. Joint Urban Remote Sens. Event*, Dubai, United Arab Emirates, Mar. 2017, pp. 1–4.
[55] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[56] C. C. Chang and C. J. Lin. (2001). *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[57] H. Kaartinen *et al.*, "An international comparison of individual tree detection and extraction using airborne laser scanning," *Remote Sens.*, vol. 4, no. 4, pp. 950–974, 2012.
[58] B. Koch, T. Kattenborn, C. Straub, and J. Vauhkonen, "Segmentation of forest to tree objects," in *Forestry Applications of Airborne Laser Scanning* (Managing Forest Ecosystems), vol. 27, M. Maltamo, E. Nasset, and J. Vauhkonen, Eds. Berlin, Germany: Springer, 2014.



WENZHI LIAO (S'10–M'14–SM'16) received the B.Sc. degree in mathematics from Hainan Normal University, Haikou, China, in 2006, the Ph.D. degree in engineering from the South China University of Technology, Guangzhou, China, in 2012, and the Ph.D. degree in computer science engineering from Ghent University, Ghent, Belgium, in 2012.

Since 2012, he has been holding a post-doctoral position at Ghent University and then a Post-Doctoral Research Fellow with the Research Foundation Flanders (FWO-Vlaanderen, Belgium). His research interests include pattern recognition, remote sensing, and image processing, and also mathematical morphology, multitask feature learning, multisensor data fusion, and hyperspectral image restoration.

Dr. Liao received the Best Paper Challenge Award from the 2013 IEEE GRSS Data Fusion Contest and the 2014 IEEE GRSS Data Fusion Contest. He is serving as an Associate Editor for the *IET Image Processing*.

FRIEKE VAN COILLIE received the M.Sc. and Ph.D. degrees in bioscience engineering (land and forest management) from Ghent University, Ghent, Belgium, in 1996 and 2003, respectively.

Since 2003, she has been associated with the Faculty of Bioscience Engineering, Laboratory of Forest Management and Spatial Information Techniques, Ghent University, as a Senior Research/Teaching Scientist. Since 2015, she has been an Appointed Professor (tenure-track), with a teaching assignment in remote sensing and GIS for vegetation management. She has authored more than 40 papers. Her research interests include the interface among remote sensing, forestry, and geo-informatics. She is a member of the organizing and scientific committees of the biannual conferences GEOBIA and ForestSAT. She has been the Associate Editor of the *International Journal of Applied Earth Observation and Geoinformation* since 2014.

LIANRU GAO (M'11–SM'18) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2007.

He is currently a Professor with the Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, CAS. He has also been a Visiting Scholar at the University of Extremadura, Cáceres, Spain, in 2014, and at the Mississippi State University, Starkville, USA, in 2017. In last 10 years, he was the PI of 10 scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China (2010–2012, 2016–2019, and 2018–2020), and the Key Research Program of the CAS (2013–2015). He has published more than 120 peer-reviewed papers, and there are 50 journal papers included by Science Citation Index (SCI). He has co-authored an academic book *Hyperspectral Image Classification and Target Detection*. His research focuses on models and algorithms for hyperspectral image processing, analysis, and applications. He obtained 17 National Invention Patents and 4 Software Copyright Registrations in China. He received the Outstanding Science and Technology Achievement Prize of the CAS in 2016 and the China National Science Fund for Excellent Young Scholars in 2017. He received the recognition of the Best Reviewer of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2015 and the Best Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2017.

LIWEI LI received the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 2008.

He is currently an Associate Professor with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences. His research interests include multisource remote sensing fusion and applications, land cover and land use mapping, and temporal remote sensing data analysis.

BING ZHANG (M'11–SM'12) received the B.S. degree in geography from Peking University, Beijing, China, and the M.S. and Ph.D. degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing.

He is currently a Full Professor and the Deputy Director of the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences (CAS), where he has been leading key scientific projects in the area of hyperspectral remote sensing for more than 20 years. He has developed five software systems in the image processing and applications. His research interests include the development of mathematical and physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas. His creative achievements were rewarded 10 important prizes from Chinese Government and special government allowances of the Chinese State Council. He received the National Science Foundation Fund for Distinguished Young Scholars of China in 2013 and the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, the highest level of awards for the CAS scholars.

He has authored more than 300 publications, including more than 170 journal papers. He has edited six books/contributed book chapters on hyperspectral image processing and subsequent applications. He is currently serving as the Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He has been serving as a Technical Committee Member of the IEEE Workshop on Hyperspectral Image and Signal Processing since 2011 and the President of Hyperspectral Remote Sensing Committee of China.

JOCELYN CHANUSSOT (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998. In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement (French National Defense Department). Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005 and an Associate Professor from 2005 to 2007 and is currently a Professor of signal and image processing. He is conducting research at the Grenoble Images Speech Signals and Automatics Laboratory. He has been a Visiting Scholar with Stanford University, USA, KTH, Sweden, and NUS, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland. From 2015 to 2017, he was a Visiting Professor with the University of California at Los Angeles, Los Angeles. His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing.

Dr. Chanussot was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008. He was a member of the IEEE Geoscience and Remote Sensing Society AdCom from 2009 to 2010, in charge of membership development. He was a member of the Institut Universitaire de France from 2012 to 2017. He was the founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010, which received the 2010 IEEE GRS-S Chapter Excellence Award. He was a co-recipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 and 2015 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award, and the IEEE GRSS 2013 Highest Impact Paper Award. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing. He was the Chair from 2009 to 2011 and the Co-Chair from 2005 to 2008 of the GRS Data Fusion Technical Committee. He was the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He was an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2005 to 2007 and *Pattern Recognition* from 2006 to 2008. Since 2007, he has been an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. In 2013, he was a Guest Editor for the PROCEEDINGS OF THE IEEE and a Guest Editor of the *IEEE Signal Processing Magazine* in 2014.