Hyperspectral and LiDAR Data Fusion Using Extinction Profiles and Deep Convolutional Neural Network

Pedram Ghamisi, *Member, IEEE,* Bernhard Höfle, and Xiao Xiang Zhu, *Senior Member, IEEE,*

***Abstract*—This paper proposes a novel framework for the fusion of hyperspectral and LiDAR-derived rasterized data using extinction profiles (EPs) and deep learning. In order to extract spatial and elevation information from both the sources, EPs that include different attributes (e.g., height, area, volume, diagonal of the bounding box, and standard deviation) are taken into account. Then, the derived features are fused via either feature stacking or graph-based feature fusion. Finally, the fused features are fed to a deep learning-based classifier (convolutional neural network with logistic regression) to ultimately produce the classification map. The proposed approach is applied to two data sets acquired in Houston, USA and Trento, Italy. Results indicate that the proposed approach can achieve accurate classification results compared to other approaches.**

***Index Terms*—Convolutional neural network, deep learning, extinction profile, graph-based feature fusion, hyperspectral, LiDAR, random forest, support vector machines.**

I. Introduction

Due to the availability of diverse remote sensors these days, it is now possible to obtain a wide variety of information from different materials on the Earth, ranging from spectral information provided by passive sensors (e.g., multispectral and hyperspectral images), to height and shape information acquired by Light Detection and Ranging (LiDAR) sensors, and texture information to amplitude and phase by Synthetic Aperture Radar (SAR). This availability makes it possible to integrate different information captured by diverse sensors to further improve object detection ability and classification performance. In spite of the rich amount of information available in such data sets, automatic interpretation of remote sensed data remains a difficult task [1].

Hyperspectral images are considered as an effective tool to define the phenomenology and spectral characteristics of the object of interest over a detailed spectral signature. LiDAR data can be taken into account to practically characterize the elevation and object height information of the scene. Many methodologies have been proposed and/or adapted to perform feature selection, feature extraction, segmentation, and classification on hyperspectral images [2–9]. In a like manner, LiDAR data have been investigated for many tasks, in particular feature detection and extraction [10–16].

However, urban scenes are usually highly complex and challenging and it is optimistic to assume that a single sensor is able to provide all the necessary information for classification and feature extraction [17]. Bearing this in mind, hyperspectral images are not applicable to effectively differentiate objects composed of the same material (i.e., objects with the same spectral characteristics). For example, roofs and roads that are made of the same material exhibit the same spectral characteristics, which makes the discrimination of such categories in the feature space a very difficult task. On the other hand, LiDAR elevation data alone cannot differentiate between objects with the same elevation that are made of different materials (e.g., roofs with the same elevation made of concrete or asphalt). In addition, the use of LiDAR data alone for complex areas, e.g., where many classes are located close to each other, is very limited compared to optical data, due to the lack of spectral information provided by this type of sensors [16, 18].

To address the above-mentioned issues and take advantage of information provided by each available sensor, the fusion of multi-sensor data can be taken into consideration. However, the automatic integration of multiple types of data is not a trivial task [19]. In addition, the use of more features extracted by different sensors, while the number of training samples is limited, may cause the so-called curse of dimensionality [9, 20, 21]. To address this issue, different feature reduction approaches, including feature extraction [22] and feature selection [23–25], can be investigated.

The joint use of hyperspectral and LiDAR data has proven to be successful for a wide variety of applications such as shadow, height, and gap-related masking techniques [26–28], above-ground biomass estimates [29], micro-climate modelling [30], quantifying riparian habitat structure [31], and fuel type mapping [32]. In addition, the joint use of LiDAR and hyperspectral data has led to higher classification accuracies compared to the use of each source individually. For instance, in [1, 14, 19, 33–35], spatial, contextual, and structural information acquired by LiDAR data has been investigated, along with spectral information captured by multispectral and hyperspectral sensors. The obtained results have shown improvement in terms of discrimination ability in forested and urban areas. In all those works, the use of LiDAR along with optical data leads to better results with respect to classification accuracies. The aforementioned works indicate that LiDAR and hyperspectral data may complement each other well and

Pedram Ghamisi and Xiao Xiang Zhu are with German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF) and Technische Universitt Mnchen (TUM), Signal Processing in Earth Observation, Munich, Germany (corresponding author, e-mail: pedram.ghamisi@dlr.de).

Bernhard Höfle is with GIScience at the Institute of Geography, Heidelberg University, Germany.

This research has been partly supported by the Alexander von Humboldt Fellowship for postdoctoral researchers, Helmholtz Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de).

by integrating those two data sets appropriately, one can make the most of the advantages of the two, while addressing the shortcomings of each of them. The sequence of research works on the joint use of LiDAR and hyperspectral data led to the 2013 Data Fusion Contest, organized by the Geoscience and Remote Sensing Society (GRSS) [1].

In [36], the concept of the attribute profile (AP) was introduced as a generalization of the morphological profile [37] to extract a multilevel characterization of an image by using a sequential application of morphological attribute filters. A comprehensive survey on APs and their capabilities for the classification of remote sensing data can be found in [9, 38]. To further improve the conceptual capability of the AP and the corresponding classification accuracies, Ghamisi *et al.* proposed extinction profiles (EPs) in 2016 [39]. EPs are based on extinction filters, which are extrema-oriented connected idempotent filters. In contrast with attribute filters, extinction filters preserve the height of the extrema kept [39]. In [39], it was shown that extinction filters are a more efficient alternative than attribute filters in terms of simplification for recognition applied to remote sensing images. This advantage leads to higher classification accuracy for EPs compared to the results obtained by APs. In addition, EPs' parameters can be set automatically, independent of the kind of the attribute being used (e.g. area, volume, ...). However, the initialization of threshold values used in APs is difficult and time-consuming. In other words, the main issue of conventional APs, the initialization of the threshold values, is addressed by EPs [39]. In [40], the concept of EPs has been generalized to extract spatial and contextual information from hyperspectral images.

Recently, classification of hyperspectral data using deep learning-based methods has attracted many researchers, due to the capability of these approaches to extract abstract features at deeper layers. More abstract features are known to be generally invariant to most local changes of the input. Deep learning is defined by the so-called "deep" neural network (DNN) architectures, commonly deeper than three layers.

Based on various architectures and activation functions, numerous classes of DNNs have been introduced, including deep belief networks (DBN) [41], deep Boltzmann machines (DBM) [42], and stacked autoencoders (SAE) [43]. The number of contributions, based on deep learning for hyperspectral image analysis is limited. In [44], a SAE-based approach was developed for hyperspectral data classification and feature extraction. In [45], a DBN-based feature extraction was developed by the same team for the classification of hyperspectral data. Although both approaches have led to acceptable classification accuracy, there is, however, a full connection between different layers. Consequently, a huge number of parameters need to be trained, which can be an undesirable factor if only a limited number of training samples is available.

Convolutional neural networks (CNNs) have gained great attention from many researchers due to their use of local connections to handle spatial dependencies. In addition, CNNs share weights, which significantly decreases the number of parameters requiring training, in comparison to other deep approaches. However, the number of parameters needed to deal with hyperspectral data is still high. In this manner, inappropriate weights may lead to getting trapped in a local minimum of the loss function. Ideally, many training samples should be available to train weights appropriately; this is an issue for hyperspectral image processing, where there is usually an imbalance between dimensionality and the number of available training samples. To partially overcome this issue, few regularization methods have been introduced to handle overfitting problems, including L2 regularization and dropout [46]. In [47], a data augmentation method called "dithering" is taken into account to further address the overfitting issue.

In this paper, a novel fusion framework is proposed for the joint classification of LiDAR and hyperspectral data. In particular, the main contributions of this paper are as follows.

1) This paper proposes a strong framework for multi-sensor data classification using EPs and graph-based multi-sensor data fusion [48][1] [35, 48]. However, the proposed methodology can be considered to be a template and therefore, different types of feature extraction and fusion approaches can be used instead of the graph-based feature fusion., and deep learning-based classification. The usefulness and generalization capability of the proposed approach have been tested on two real data sets with different land-covers. To this end, the first data set, Houston data, is taken over an urban area, while the second data set is taken over a rural area in Trento, Italy.
2) To the best of our knowledge, this paper investigates a deep learning-based approach for the classification of multisensor data, LiDAR and hyperspectral, for the first time in the remote sensing community.
3) The concept of EPs has successfully been investigated for the classification of panchromatic [39] and hyperspectral data [40] so far. This paper also investigates the ability of the EPs to extract useful information from LiDAR images.

The rest of the paper is organized as follows: Section II is devoted to the methodology. Section III presents experimental results on two well-known data sets. Section IV wraps up the paper by providing the main concluding remarks.

II. METHODOLOGY

This paper considers two strategies to fuse elevation, spectral, and spatial information of LiDAR rasterized data and hyperspectral images. Figs. 1 and 2 show the proposed fusion strategies, strategy 1 and strategy 2, respectively. In summary:

1) Strategy 1 (Fig. 1): In this framework, nonparametric-weighted feature extraction (NWFE), an extended multi-extinction profile (EMEP), and a multi-extinction profile (MEP) are applied to the LiDAR and hyperspectral images to extract spectral, spatial, and elevation information. Finally, all extracted features are concatenated into a stacked vector and fed to a classifier (RF or

[1]Here, we only used graph-based feature fusion since its performance has already been proven to be successful for the fusion of LiDAR and hyperspectral data.

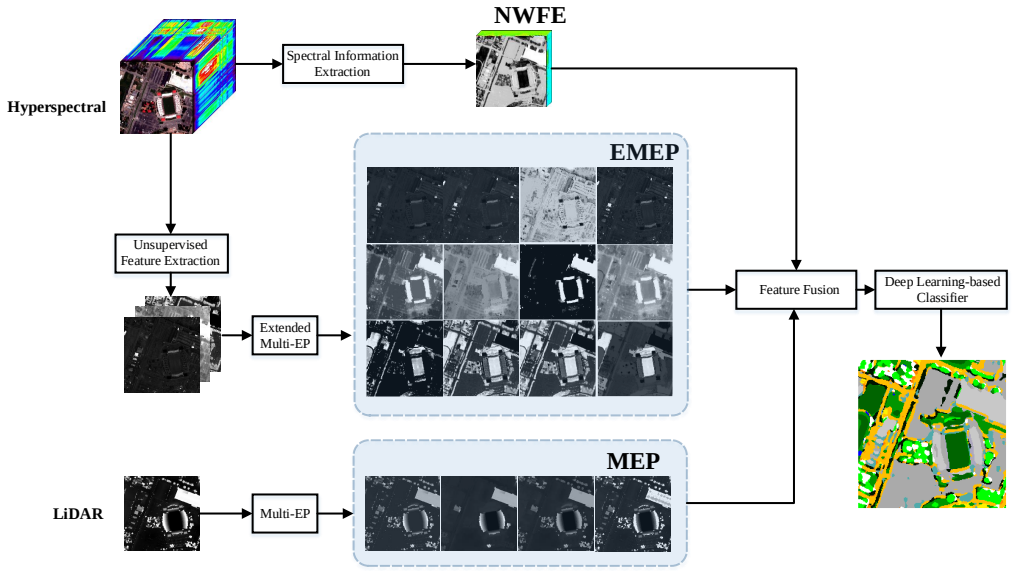

Fig. 1. Strategy 1: The architecture of the proposed method using feature stacking as the fusion step.

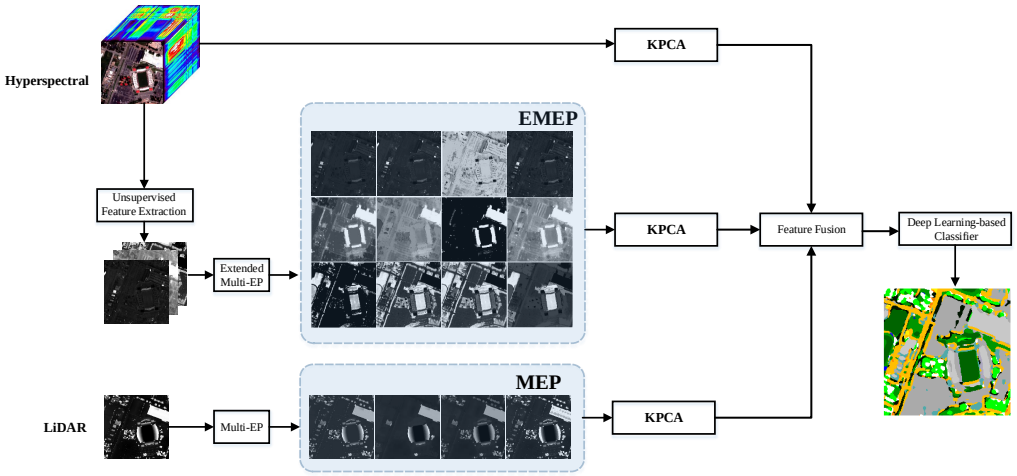

Fig. 2. Strategy 2: The architecture of the proposed method using graph-based feature fusion as the fusion step.

CNN-based classification approach) to produce the final classification map.

2) Strategy 2 (Fig. 2): In this framework, EMEP is applied to the input hyperspectral data to extract spatial information. In parallel, MEP is applied to the LiDAR-derived Digital Surface Model (DSM) image to extract elevation information. In order to normalize the number of spectral, spatial, and elevation features, kernel principal component analysis (KPCA) is separately applied to the input data, the output of EMEP, and the output of MEP. Extracted features are fused using the graph-based feature fusion (GBFF) and fed to a classifier [random forest (RF) [49] or CNN-based classification approach] to produce the final classification map. RF has already been shown to be effective in terms of classification accuracy and efficient in terms of CPU processing time, when it is performed on the features extracted by MPs and APs, and frequently outperforms well-known classifiers in the hyperspectral community, such as support vector machines (SVMs), on such features [9, 38]. We used RF here to evaluate and compare the performance of the proposed method with the RF.

For both strategies, we tried to feed the spectral, spatial, and elevation information to the final classification approach. To this end, EPs automatically generate spatial and elevation features from the DSM (a LiDAR-derived feature) and the first ICs (hyperspectral derived features). It is not recommended to perform approaches such as MPs, APs, and EPs directly on the whole hyperspectral data set, as they produce many redundant features due to the high redundancy available between hyperspectral bands. Instead, they have almost often been performed on a few features extracted by a feature extraction approach (here, ICA). Below, a detailed description of the main building blocks of the proposed strategies and the reason for their use is provided.

A. Extinction Profile (EP)

1) Extinction Filters: Ghamisi *et al.* [39] proposed the concept of extinction profiles (EPs), based on a set of connected filters, called extinction filters, which can preserve relevant image extrema. Relevance here can be measured using the concept of the extinction value defined by Vachier [50]. Let $Max(f) = \{M_1, M_2, ..., M_N\}$ be the regional maxima of the image $\mathbf{F}$. For each regional maxima, M_i, there is an extinction value ϵ_i corresponding to the increasing attribute being analyzed. For the input gray-scale image $\mathbf{F}$, the extinction filter preserves the n maxima with the highest extinction values, which can be shown as follows:

$$\mathrm{EF}^n(\mathbf{F}) = R_g^{\delta}(\mathbf{F}), \tag{1}$$

where $R_g^{\delta}(\mathbf{F})$ denotes the reconstruction by dilation [51] of the mask image, which is given as follows:

$$g = \max_{i=1}^{n}\{M_i'\}, \tag{2}$$

where $\max$ is the pixel-wise maximum operation. The term M_1' is the maximum with the highest extinction value, M_2' has the second highest extinction value, and so on. By construction, the transformation that defines any regional extrema of an image with the corresponding extinction value defines the concept of a granulometric operation [52], which is a family of opening and closing operators of increasing size.

The efficient implementation of the extinction filter is based on max-tree data representation [53]. After constructing the max-tree, the n maxima (max-tree leaves) with the highest extinction values for the corresponding attribute are chosen, while all other max-tree nodes that are not in the paths from these leaves to the root are pruned.

In [39], it was shown that extinction filters act more efficiently than attribute filters with respect to simplification for recognition of remote sensing panchromatic images, since they can preserve more regions and correspondences found by affine region detectors. Another advantage of extinction filters over attribute filters is that it is easier to set the parameters of the extinction filters than those of attribute filters. The main reason is that they are independent from the kind of attribute being used (e.g., area, volume,...), since they are based on the number of extrema. In contrast, the thresholds used by attribute filters vary greatly according to the attribute being used as well as the data set being analyzed. Therefore, the thresholds are more difficult to set.

2) Extended Multi-Extinction Profile (EMEP): The main idea behind using EPs is to apply several extinction filters with progressively higher threshold values to appropriately extract and model the spatial information of the adjacent pixels. In more detail, the EP is constructed by performing a sequence of thinning and thickening transformations defined with a sequence of progressively stricter criteria. The EP for the input gray scale image, $\mathbf{F}$, can be presented as:

$$\begin{aligned}\mathrm{EP}(\mathbf{F}) = \\ \{\underbrace{\phi^{P_{\lambda_L}}(\mathbf{F}), \phi^{P_{\lambda_{L-1}}}(\mathbf{F}), \ldots, \phi^{P_{\lambda_1}}(\mathbf{F})}_{\text{thickening profile}}, \mathbf{F}, \\ \underbrace{\gamma^{P_{\lambda_1}}(\mathbf{F}), \ldots, \gamma^{P_{\lambda_{L-1}}}(\mathbf{F}), \gamma^{P_{\lambda_L}}(\mathbf{F})}_{\text{thinning profile}}\},\end{aligned} \tag{3}$$

where $P_\lambda : \{P_{\lambda_i}\}$ $(i = 1, \ldots, L)$ is a set of L ordered predicates (i.e., $P_{\lambda_i} \subseteq P_{\lambda_k}$, $i \leq k$). For EPs, the number of extrema can be considered as the predicates. The terms ϕ and γ are thickening and thinning transformations.

The EP, as presented here, only works on a gray-scale image. To further generalize the concept of the EP to hyperspectral data, one possible way is to perform a feature reduction approach, such as PCA or independent component analysis (ICA), on the input data and then, apply EPs to the most informative features [9]. This approach is based on the reduction of the dimensionality of the data from $E \subseteq \mathbf{Z}^n$ to $E' \subseteq \mathbf{Z}^m$ $(m \leq n)$ with a generic transformation $\Psi : E \to E'$ carried out on an input image $\mathbf{F}$ (i.e., $\mathbf{Q} = \Psi(\mathbf{F})$). Then, the EP can be performed on the most informative features $\mathbf{Q}_i$ $(i = 1, \ldots, m)$ of the extracted features, which can mathematically be given as:

$$\mathrm{EEP}(\mathbf{Q}) = \{\mathrm{EP}(\mathbf{Q}_1), \mathrm{EP}(\mathbf{Q}_2), \ldots, \mathrm{EP}(\mathbf{Q}_m)\}. \tag{4}$$

In contrast to MPs that are only able to model the size and structure of different objects, EPs are more flexible and can be of any type. In this way, the extended multi-EP (EMEP) concatenates different EEPs (e.g., area, height, volume, diagonal of bounding box, and standard deviation on different extracted features) into a single stacked vector, which can be mathematically defined as follows:

$$\mathrm{EMEP} = \{\mathrm{EP}_{a_1}, \mathrm{EP}_{a_2}, ..., \mathrm{EP}_{a_w}\}, \tag{5}$$

where a_k, $k = \{1, ..., w\}$ denotes different types of attributes. Since different extinction attributes can extract complementary spatial information, the EMEP has a greater ability to extract spatial information than a single EP.

It should be noted that the EMEP demands almost the same computational time as a single EP, since the most time demanding step is to produce the max-tree and min-tree, which are computed only once for each gray-scale image.

In our experiments on the LiDAR-derived image, since there is only one image available, we use the term multi-EP (MEP) for the situation when different types of EPs are applied to the LiDAR image.

3) EEP *Computational Complexity Analysis:* It is easy to obtain the fact that the computational complexity of the EEP is m times the complexity of computing EP, in which m is the number of informative features retained after performing ICA or PCA.

The most time-consuming part is the construction of the max-tree and min-tree required to compute the thickening and thinning profiles. The complexity for a generic floating point structure is $O(n\log n)$, where n is the number of image pixels.

TABLE I
COMPLEXITY ANALYSIS OF THE EEP. THE PARAMETER "s" REFERS TO THE NUMBER OF THRESHOLD VALUES IN THE PROFILE. THE PARAMETER "m" REPRESENTS THE NUMBER OF INFORMATIVE FEATURES KEPT AFTER PERFORMING A FEATURE REDUCTION APPROACH.

Operation	Complexity	# Occurrence
Max-tree construction	$O(n\log\, n)$	$2m$
Attribute computation	$O(n)$	$2m$
Extinction values computation	$O(m)$	$2m$
Filtering	$O(n)$	$2ms$

For a complete analysis of the max-tree construction complexity for different data types and different implementations, refer to [54].

In our implementation, we use the array-based node-oriented max-tree representation proposed in [55]. This representation is very flexible, and for some attributes, such as height, it reduces the computational complexity from $O(n)$ to $O(m)$, where m is the number of max-tree nodes. The structure is also suitable for parallel processing of the max-tree. Table I demonstrates the computational complexities of different steps in the EEP. For detailed information about the complexities of the max-tree construction, attributes computation and filtering, see [54, 55].

B. Convolutional Neural Network (CNN)

Compared to other deep approaches, CNNs [56] take advantage of local connections and shared weights. CNNs exploit local correlations using local connectivity between the neurons of near layers. In CNNs, some connections between neurons, which share the same weights and biases, are replicated across the entire layer. Fig. 3 demonstrates an example of the CNN-based classification. As can be seen, the CNN consists of several convolutional and pooling layers that construct a deep network. In order to use the CNN for classification, a fully connected logistic regression (LR) layer can be considered at the end of the network. A convolutional layer is as follows:

$$x_j^l = f\left(\sum_{i=1}^{D} x_i^{l-1} * k_{ij}^l + b_j^l\right),$$

where x_i^{l-1} is the *i*-th feature map of the *(l-1)*-th layer, x_j^l is the *j*-th feature map of the current *(l)*-th layer, and *D* is the number of input feature maps. The k_{ij}^l and b_j^l are the trainable parameters in the convolutional layer. The function $f(.)$ is a nonlinear function and $*$ is the convolution operation.

In addition to the convolutional layer, one can also use pooling in a network. The main advantage of pooling is that such approaches can extract invariant features by reducing the resolution of feature maps. As shown in Fig. 3, each pooling layer is connected to the previous convolutional layer, and combines a small $N \times 1$ patch of the convolution layer. The most common pooling technique is max pooling [57].[2]

Due to the high dimensionality of hyperspectral data, a network can be forced to overfitting. To handle this issue to some extent, rectified linear units (ReLU), dropout layers, and dithering [47] can be taken into account. For detailed information about CNNs and their design, please see [58].

The output of CNN is classified using an LR, which employs soft-max as its output-layer activation. Soft-max ensures that the activation of each output unit sums to one. Therefore, the output can be seen as a set of conditional probabilities. LR can be considered to be a single layer neural network and, as a result, it can be merged with the CNN to form a CNN+LR deep classifier. In this manner, the size of the output layer should be equal to the number of classes.

Both the EPs and the CNN are considered to extract meaningful features from the input data. Here, it should be noted that, although deep models extract abstract features in their deep layers in particular, the deep learning model can be considered to be a feature refiner, i.e., mapping the input low-level feature to a mid/high-level one.

C. Data Fusion

1) Feature Stacking: In regards to Strategy 1, feature stacking is used for feature fusion. Feature stacking is a simple approach to integrating extracted features from LiDAR and hyperspectral images. In this manner, let $\mathbf{X}^{\text{Spe}}$ denote the input hyperspectral data, and let $\mathbf{X}^{\text{Spa}}$ be the output of the EPs on the first informative independent components of the hyperspectral data, which can extract and model spatial information of the hyperspectral data. Let $\mathbf{X}^{\text{Ele}}$ denote the features obtained by performing MEP on the LiDAR image, which can extract elevation information. Unsupervised feature extraction approaches, such as ICA and PCA, do not consider the class-specific information of hyperspectral data, which can be provided by training samples. To efficiently extract spectral information while decreasing the dimensionality, supervised feature extraction approaches, such as NWFE, can be taken into account [59, 60]. In this case, let $\mathbf{X}^{\text{NWFE}}$ denote the features extracted by the NWFE. The feature stacking approach simply concatenates the features, i.e., $\mathbf{X}^{\text{Sta}} = \left[\mathbf{X}^{\text{Spa}}; \mathbf{X}^{\text{Ele}}; \mathbf{X}^{\text{NWFE}}\right]$. The main shortcoming of such approaches is that they increase dimensionality in the feature space, which may cause the Hughes phenomenon [20]. This issue might dramatically downgrade the classification accuracy of classifiers, which cannot handle high dimensionality with a limited number of training samples.

2) Graph-Based Feature Fusion (GBFF): In regard to Strategy 2, a GBFF developed by Liao *et al.* [48] is used for the fusion of spectral, spatial, and elevation features. The outputs of different steps may have different dimensionalities and characteristics, detailed as follows:

1) For example, the output of EPs on the first three independent components produces 213 features (i.e., 71 features for each independent component, including 14 features for the height attribute, 14 features for the area attribute, 14 features for the volume attribute, 14 features for the diagonal of the bounding box attribute, and 14 features for the standard deviation attribute, and the independent component should also be included to make a complete profile). These features extract the spatial information of the scene and model different attributes. Let $\mathbf{X}^{\text{Spa}}$ denote the spatial features.

[2]We have used the MatConvNet library for the implementation of the CNN-based classification method utilized in http://www.vlfeat.org/matconvnet/.

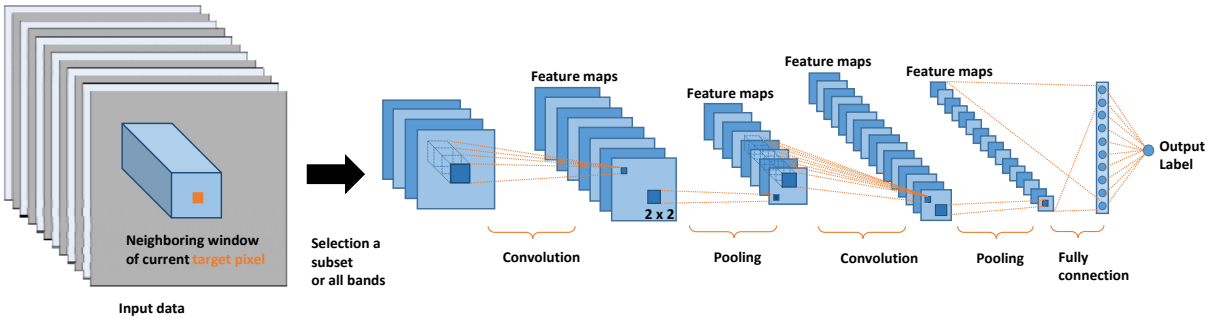

Fig. 3. A general example of a CNN network. For strategy 1, a concatenation of NWFE, EMEP, and MEP is used as the input for the CNN deep network, while for strategy 2, the output of the graph-based feature fusion is fed to the CNN deep network as the input.

2) Hyperspectral data sets contain detailed spectral information. For example, in the case of the Houston data, the number of spectral features is 144. Let $\mathbf{X}^{\text{Spe}}$ denote the spectral features.
3) By performing the MEP on the LiDAR image, one may obtain several features presenting elevation information of the LiDAR derived data. (i.e., 71 features for the LiDAR image). Let $\mathbf{X}^{\text{Ele}}$ denote the elevation features.

In order to fuse the features described above, the number of dimensionalities should first be normalized in order to put the same weight on each type of the features and reduce the computational cost and noise throughout the feature space [1]. In [1], Kernel PCA [61] was suggested as an effective tool to reduce the dimensionality of each type of features, separately, since it can represent a higher-order complex and nonlinear distribution in a fewer number of dimensions, which is helpful against the Hughes phenomenon and high computational cost. As suggested in [1], the normalized dimension of each type of features is set to the smallest dimensions of the above-mentioned features. For example, for the Houston data, this value is set to 71 (d_1 =71).

Let $\mathbf{F}^{\text{Spe}} = \left\{\mathbf{F}_i^{\text{Spe}}\right\}_{i=1}^{d_1}$, $\mathbf{F}^{\text{Spa}} = \left\{\mathbf{F}_i^{\text{Spa}}\right\}_{i=1}^{d_1}$, and $\mathbf{F}^{\text{Ele}} = \left\{\mathbf{F}_i^{\text{Ele}}\right\}_{i=1}^{d_1}$ represent the spectral, spatial, and elevation features after normalization to the same number of dimensions, respectively, while $\mathbf{F}_i^{\text{Spe}} \in \Re^{d_1}$, $\mathbf{F}_i^{\text{Spa}} \in \Re^{d_1}$, and $\mathbf{F}_i^{\text{Ele}} \in \Re^{d_1}$ are the normalized spectral, spatial and elevation features, respectively. Let $\mathbf{F}^{\text{Sta}} = \left[\mathbf{F}^{\text{Spe}}; \mathbf{F}^{\text{Spa}}; \mathbf{F}^{\text{Ele}}\right]$ and $\mathbf{F}_i^{\text{Sta}} = \left[\mathbf{F}_i^{\text{Spe}}; \mathbf{F}_i^{\text{Spa}}; \mathbf{F}_i^{\text{Ele}}\right] \in \Re^{3(d_1)}$ denote the vector stacking of the spectral, spatial, and elevation features. Finally, let $\{\mathbf{Z}_i\}_{i=1}^{n}$ and $\mathbf{Z}_i \in \Re^{d_2}$ represents the fusion features with dimensionality of d_2 with $d_2 \leq 3(d_1)$.

The main aim of the graph-based feature fusion is to seek a transformation matrix, $\mathbf{w} \in \Re^{3(d_1)\times d_2}$, which can perform both dimensionality reduction and feature fusion in such a way that $\mathbf{Z}_i = \mathbf{w}^T\mathbf{F}_i$, where $\mathbf{F}_i$ can be set to $\mathbf{F}_i^{Sta}$. The transformation matrix $\mathbf{w}$ can reduce dimensionality and fuse features at the same time, while it preserves local neighborhood information and detects manifolds embedded in the original feature space [48]. To do so, the following approach can be considered to seek an appropriate transformation matrix $\mathbf{w}$:

$$\arg \min_{\mathbf{w}\in\Re^{3d_1\times d_2}} \left(\sum_{i,j=1}^{n} \left\| \mathbf{w}^T\mathbf{F}_i - \mathbf{w}^T\mathbf{F}_j \right\|^2 \mathbf{A}_{ij} \right),$$

where matrix $\mathbf{A}$ is denoted as the edge of the graph $\mathbf{G} = (\mathbf{F}, \mathbf{A})$.

III. Experimental Results

A. Data Description

1) Houston Data: The data is composed of a hyperspectral image and a LiDAR-derived digital surface model (DSM). This data set was distributed for the 2013 GRSS data fusion contest. The hyperspectral data was acquired by the Compact Airborne Spectrographic Imager (CASI) over the University of Houston campus and the neighboring urban area on June 23, 2012. The LiDAR data was acquired on June 22, 2012. The data sets were collected by the NSF-funded Center for Airborne Laser Mapping (NCALM). The size of the data is 349×1905 pixels with the spatial resolution of 2.5m. The hyperspectral data set consists of 144 spectral bands ranging from 0.38 to 1.05μm. The 15 classes of interests are: Grass Healthy, Grass Stressed, Grass Synthetic, Tree, Soil, Water, Residential, Commercial, Road, Highway, Railway, Parking Lot 1, Parking Lot 2, Tennis Court, and Running Track. The "Parking Lot 1" includes parking garages at the ground level and also in elevated areas, while "Parking Lot 2" corresponds to parked vehicles. Fig. 4 shows a color composite representation of the hyperspectral data and the corresponding training and test samples. Table II gives information about the number of training and test samples for different classes of interests.

Cloud shadows in the hyperspectral data were detected using thresholding of illumination distributions calculated by the spectra. Relatively small structures in the thresholded illumination map were removed based on the assumption that cloud shadows are larger than structures on the ground.[3]

2) Trento Data: The second data set is a subset of larger data captured over a rural area south of the city of Trento, Italy. The subset used in the experiments is of 600 by 166 pixels. The LiDAR DSM data was acquired by the Optech ALTM

[3]The enhanced data is provided by Prof. Naoto Yokoya from Technical University of Munich (TUM).

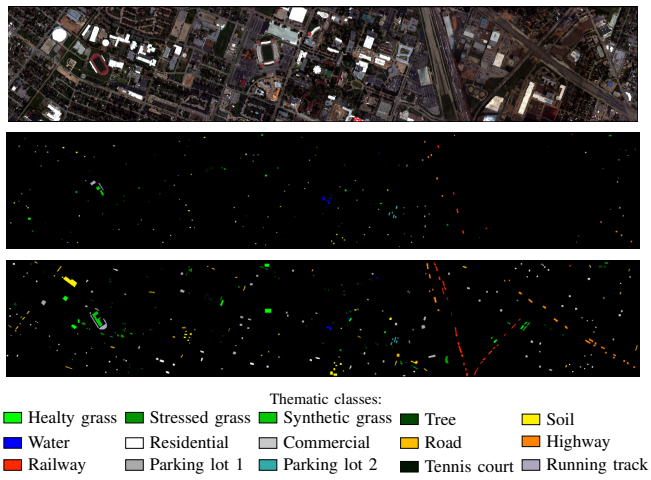

Fig. 4. Houston - From top to bottom: A color composite representation of the hyperspectral data using bands 70, 50, and 20, as R, G, and B, respectively; training samples; test samples; and legend of different classes.

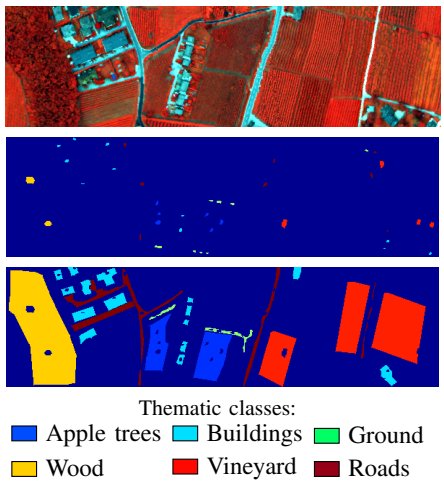

Fig. 5. Trento - From top to bottom: A color composite representation of the hyperspectral data using bands 40, 20, and 10, as R, G, and B, respectively; Training samples; Test samples; and legend of different classes.

TABLE II
HOUSTON: NUMBER OF TRAINING AND TEST SAMPLES.

Class		Number of Samples	
No	Name	Training	Test
1	Grass Healthy	198	1053
2	Grass Stressed	190	1064
3	Grass Synthetic	192	505
4	Tree	188	1056
5	Soil	186	1056
6	Water	182	143
7	Residential	196	1072
8	Commercial	191	1053
9	Road	193	1059
10	Highway	191	1036
11	Railway	181	1054
12	Parking Lot 1	192	1041
13	Parking Lot 2	184	285
14	Tennis Court	181	247
15	Running Track	187	473
Total		2,832	12,197

3100EA sensor and the hyperspectral data captured by the AISA Eagle sensor, all with the spatial resolution of $1m$. The hyperspectral data consists of 63 bands ranging from 402.89 to 989.09nm, where the spectral resolution is 9.2nm. The spatial resolution of this data set is $1m$. For this data set, six classes of interests were extracted, including Building, Woods, Apple Trees, Roads, Vineyard, and Ground. Fig. 5 shows a color composite representation of the hyperspectral data and the corresponding training and test samples. Table III gives information about the number of training and test samples for different classes of interests.

B. Algorithm Setup and Discussion

For NWFE, the first features with cumulative eigenvalues above 99% are automatically retained. To this end, 6 and 17 features have been extracted from the Trento and Houston data, respectively.

For the RF, the number of trees is set to 300. The number of the prediction variable is set approximately to the square root of the number of input bands.

TABLE III
TRENTO: NUMBER OF TRAINING AND TEST SAMPLES.

Class		Number of Samples	
No	Name	Training	Test
1	Apple trees	129	3905
2	Buildings	125	2778
3	Ground	105	374
4	Woods	154	8969
5	Vineyard	184	10317
6	Roads	122	3252
Total		819	29595

Fig. 6 and Table IV provide information about the structure of the CNN network. Furthermore, the following points have been taken into account to design a proper CNN network:

1) In this work, 90% of the training samples were used to train weights and biases, while the rest are validation samples to guide the design of a proper architecture in order to avoid overfiting. Please note that the validation samples are different from the test samples. The validation samples have been extracted directly from the training set.
2) A dynamic learning rate is taken into account. To do so, the whole training period is divided into five stages, starting from a relatively high learning rate of 0.005, and decreasing by a half for each subsequent stage. This setting provides a fast-descend-of-loss function at the beginning, while the gradually decreasing rate can ensure a small but consistent progress. After reaching a certain stage of training, a high rate might not be efficient anymore, since it might cause oversteps leading to a higher loss.
3) In this paper, the input hyperspectral data sets were normalized in the range of $[0\ \ 1]$. In order to extract sufficient spatial information for the pixel to be classified, a large window with the size of 27×27 pixels was

TABLE IV
THE ARCHITECTURE OF THE CNN.

Number	Convolution	ReLU	Pooling	Dropout
1	$4 \times 4 \times 32$	Yes	2×2	No
2	$5 \times 5 \times 64$	Yes	2×2	50%
3	$4 \times 4 \times 128$	Yes	No	50%

INPUT	**[27 × 27 × 10]**
CONV-1	kernel size: 4 × 4 × 10 kernel #: 32 weights: (4 × 4 × 10) × 32 + 32 (bias)
Feature Map-1	**[24 × 24 × 32]**
POOL-1	size: 2 × 2
Feature Map-2	**[12 × 12 × 32]**
CONV-2	kernel size: 5 × 5 × 32 kernel #: 64 weights: (5 × 5 × 32) × 64 + 64 (bias)
Feature Map-3	**[8 × 8 × 64]**
POOL-2	size: 2 × 2
Feature. Map-4	**[4 × 4 × 64]**
CONV-3	kernel size: 4 × 4 × 64 kernel #: 128 weights: (4 × 4 × 64) × 128 + 128 (bias)
Feature Map-5	**[1 × 1 × 128]**
Full Connection	kernel size: 1 × 1 × 128 kernel #: 9 weights: (1 × 1 × 128) × 16 + 16 (bias)
Feature Map-6	**[1 × 1 × 16]**
Softmaxloss	
Output	probability vector: [1 × 16]

Fig. 6. Detailed information about the network considered for CNN and SICNN.

considered. Since the studied areas are of small sizes, only three convolution layers and two pooling layers have been used.

4) For the training step, a mini batch with a size of 32 was taken into account. For relatively small training samples, as in our case, this could allow the training step to perform more frequent parameter updates and achieve faster convergence in practice.
5) In order to preserve the borders of different features through the convolutional process, the original image is padded with an extra artificial border, which mirrors the original border.

In terms of EPs, the following points have been taken into account:

1) In order to generate the EP for area, volume, and diagonal of the bounding box, the values of n used to generate the profile are automatically given by $\lfloor \alpha^j \rfloor$, where $j = 0, 1, ..., s-1$ and s shows the number of thresholds. The aforementioned equation was obtained experimentally. In this equation, the larger the α, the larger the differences between consecutive images in the profile. The smaller the α, the fewer extrema there will be, where most of the image information is usually present [53]. As recommended in [39], the appropriate α value can be chosen between 2 and 5. Here, α, and s are set to 3 and 7, respectively.
2) In order to generate the EP for height and standard deviation, the threshold values were adjusted with respect to the maximum value of each attribute, disregarding extreme values such as the root node, which usually has a much higher attribute than the other nodes [39]. Then, the maximum value is split up into seven equidistant parts.
3) The size of the EPs is $2s + 1$, since the original image should also be included in the profile. The profiles were computed using the 4-connected connectivity rule.

Fig. 7 shows a few representative features produced by EPs on the LiDAR image using area, volume, diagonal of the bounding box, height, and standard deviation attributes. As can be seen, different extinction attributes extracts different spatial information, which can be suitable for classification accuracies.

For the sake of simplicity, the following names are used in the experimental part: **LiDAR**, **Hyper**, and **LiDAR+Hyper** show the classification accuracies of LiDAR, hyperspectral, and their stack, respectively. $\mathbf{EP}_{\text{lidar}}$ and $\mathbf{EP}_{\text{hyper}}$ show the classification accuracies of EPs applied to LiDAR, and hyperspectral data. $\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ refers to the classification accuracies of EPs applied to the stack of LiDAR and hyperspectral. **GBFF** and **Stack** show the classification accuracies of the proposed method using GBFF and stacking as the fusion step.

C. Discussion of the Houston Data

1) RF-based approaches: Table V shows the classification accuracies obtained by different approaches using RF. The classification accuracies obtained by **LiDAR+Hyper** improves both the classification results obtained by the individual use of **LiDAR** and **Hyper**, which confirms that LiDAR and hyperspectral data are appropriate complements for each other in terms of classification accuracies. The use of EPs can significantly improve kappa, overall accuracy (OA), and average accuracy (AA), since the EPs can efficiently extract spatial information and model the shape and size of different objects, which are helpful to precisely differentiate different classes of interest. For example, $\mathbf{EP}_{\text{lidar}}$ significantly improves the classification accuracy of **LiDAR** by almost 42% in terms of the OA, which confirms the capability of the EP in terms of information extraction from LiDAR-derived rasterized data. The best results were obtained by the proposed approach using feature stacking. In this context, **Stack** improves **GBFF** by almost 1.5% in terms of OA. The main reason for this improvement is that RF is insensitive to noise in the training labels and can handle high dimensional data with even a limited number of training samples. The **GBFF** approach reduces dimensionality and discards some information, while feature stacking increases the dimensionality and keeps all the information. Due to the robustness of the RF, a better classification result can be obtained by the proposed approach using feature stacking. In terms of class-specific accuracies, **Stack** also improves the class-specific accuracies of most classes compared to $\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$. The main reason is that EP is performed on the first ICs generated by ICA. ICA does not consider class information provided by training samples and therefore, it cannot extract spectral information properly. However, in **Stack**, a few extra features produced by NWFE are also considered, which can extract spectral information with respect to class discriminant information provided by training samples. The only exceptions are classes Roads, Highways, and Parking lots, where the classes are made of almost exactly the same materials and therefore have almost the same spectral characteristics. As a result, spatial

and elevation information can be more helpful to differentiate these classes than spectral information.

2) CNN-based approaches: Table VI shows the classification accuracies obtained by different approaches using CNN as the classifier. The use of EP significantly improves the classification accuracies of both LiDAR and hyperspectral data due to the great ability of the EP in terms of simplification for recognition. As can be seen, the proposed approach using graph-based feature fusion provides the best results in terms of classification accuracies. The main reason is that CNN, in contrast with RF, is not that efficient at handling high dimensional data when the number of training samples is insufficient. The graph-based approach performs the fusion step while it reduces the dimensionality. However, the feature stacking-based approach increases dimensionality by concatenating elevation, spectral, and spatial features. This is the main reason why **Graph** outperforms **Stack** when CNN is chosen for the classification step. In addition, **Graph** preserves local neighborhood information in the projected lower dimensional feature space, while it detects the manifold embedded in the original high-dimensional input data.

In table VI, $\mathbf{EP}_{\text{hyper}}$ shows the best class-specific accuracy for the category highways. The reason is that for this particular class, the LiDAR-derived features complicate the distribution of classes in the feature space as they consider elevation information to distinguish different classes. For the Houston data, the elevation of this particular class changes along the highways. Therefore, the consideration of EPs on the hyperspectral data set alone can lead to the best classification performance for the highways, as it only considers spectral information.

Fig. 8 demonstrates a few classification maps obtained by applying different approaches to the Houston data. In this manner, the outputs of RF on (a) hyperspectral data, (b) the stack of LiDAR and hyperspectral data, and (c) the proposed approach using feature stacking; and the outputs of CNN-based classification on (d) hyperspectral data, (e) the stack of LiDAR and hyperspectral data, and (f) the proposed approach using GBFF are demonstrated. As can be seen, the consideration of EPs can produce more homogeneous classification maps than **Hyper** and the stack of LiDAR and hyperspectral. Although the cloud shadow was removed from the original data, when ICA is performed to the enhanced hyperspectral data to produce base features for MEEPs, the shadow effect is partially appeared on the second IC. This is the reason that the cloud shadow slightly downgrades the quality of the classification maps obtained by the proposed approach.

D. Discussion on the Trento Data

With respect to Tables VII, and Table VIII, one can simply notice that in all cases, $\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$, **Graph**, and **Stack** could lead to the highest classification accuracies. In addition, the consideration of the spatial and elevation information can significantly boost the performance of using **LiDAR** and **Hyper** in terms of classification accuracies, when these sources have been considered separately. Same as the Houston data, the use of EPs can considerably improve the classification accuracy of **LiDAR**. In this context, the amount of improvement by RF, and CNN are almost 41%, and 35%, respectively. In all cases, the **LiDAR+Hyper** can boost the performance of either **LiDAR** or **Hyper** in terms of classification accuracy, which proves that the consideration of elevation information along with spectral information are suitable in terms of obtaining accurate classification maps. In terms of CNN, **Graph** slightly improves **Stack** in terms of classification accuracies due to the fact that **Graph** reduces dimensionality while fusing different features, which is suitable from the stand point of classification accuracy for CNN. Fig. 9 demonstrates a few classification maps obtained by different approaches on the Trento data.

E. Comparison with the Literature

In this section, we briefly compare the proposed approach with the literature. For the Trento data set, the proposed approach improved the methodologies published in [19, 62] in terms of classification accuracies. Indeed, the CNN considered in our methodology provides the highest classification accuracies. In terms of SVM and RF, our fusion framework also leads to higher classification accuracy than the ones reported in [19, 62]. This improvement might be due to the use of EPs instead of APs in the proposed approach. With respect to the study published in [39], EPs are more powerful and efficient than APs in terms of simplification for recognition applied to remote sensing images and they can preserve the height of the retained extrema.

For the Houston data, with respect to the outcome of the 2013 Fusion Contest,[4] the proposed approach provides acceptable classification accuracies. However, it is important to note that most approaches investigated in the contest had been specifically developed for the classification of the Houston data, while they include several overheads, pre-processing, and postprocessing approaches to further improve the eventual classification accuracy. The consideration of these pre- and post-processing approaches can also be an interesting research line to further improve the obtained classification accuracies. However, in this paper, we have tried to propose a scheme that is also applicable to other data sets composed of coregistered hyperspectral and LiDAR images. In other words, the main objective of the paper is to propose an efficient fusion framework that is capable of handling different coregistered LiDAR and hyperspectral images by preserving the generalization capability of the proposed approach, while achieving the highest classification accuracy on one specific data set is not expected. In [63], a fusion framework using multiple feature learning was developed for the Houston data, whose results are comparable to the ones obtained by the proposed approach. The proposed approach can slightly improve the results of the work in [63] when no postprocessing is taken into account. However, when Markov random field (MRF) is used as the post-processing approach, the classification result of [63] will be slightly better than the proposed approach, in which no postprocessing is used. This fact encourages us to consider

[4]http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-classification-contest-results/

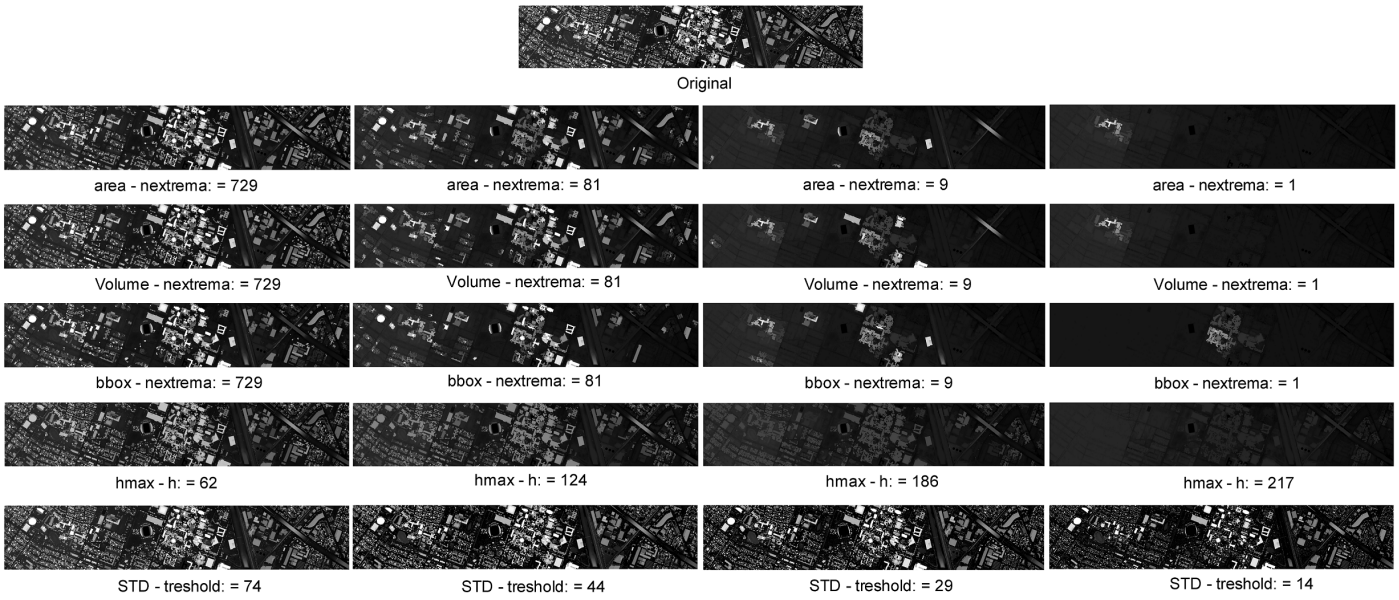

Fig. 7. A few representative features produced by EPs on the LiDAR-derived image file using area, volume, diagonal of the bounding box, height, and standard deviation attributes.

Fig. 8. Classification maps for Houston data: The outputs of RF on (a) hyperspectral data, (b) the stack of LiDAR and hyperspectral data, and (c) the proposed approach using feature stacking; the outputs of CNN-based classification on (d) hyperspectral data, (e) the stack of LiDAR and hyperspectral data, and (f) the proposed approach using GBFF.

an approach like the hidden MRF proposed in [2] to further improve the classification accuracy of the proposed approach in future.

IV. Conclusion

This paper proposes a fusion approach for the spectral-spatial classification of LiDAR and hyperspectral data using extinction profiles and convolutional neural networks. In this work, the concept of the extinction profile has been used for spatial, and elevation information extraction from both LiDAR and hyperspectral data. The spectral, spatial, and elevation features were then fused using either feature stacking or graph-based feature fusion. Finally, the features were classified using a few advanced approaches, such as RF and CNN-based classification. The principal conclusions are as follows

1) EPs can significantly improve the classification accuracy of both LiDAR and hyperspectral data. In this paper, the usefulness of the EP for spatial and elevation information extraction from LiDAR data has been investigated for the first time in the remote sensing community. Results indicate that promising results can be obtained using the EP on LiDAR data, without involving any information from hyperspectral data.
2) In this work, feature fusion was a better option than graph-based feature fusion where RF is considered for the classification step. On the other hand, a further

TABLE V
RF HOUSTON: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT APPROACHES USING RF. THE METRICS AA AN OA ARE REPORTED IN PERCENTAGE. KAPPA COEFFICIENT IS OF NO UNITS. THE BEST RESULT IS SHOWN IN BOLD. **LIDAR**, **HYPER**, AND **LIDAR+HYPER** SHOW THE CLASSIFICATION ACCURACIES OF LIDAR, HYPERSPECTRAL, AND THEIR STACK, RESPECTIVELY. $\mathbf{EP}_{\text{lidar}}$, $\mathbf{EP}_{\text{hyper}}$, AND $\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ SHOW THE CLASSIFICATION ACCURACIES OF EPS APPLIED TO LIDAR, HYPERSPECTRAL, AND THE STACK OF MEP AND EMEP, RESPECTIVELY. **GBFF** AND **STACK** SHOW THE CLASSIFICATION ACCURACIES OF THE PROPOSED METHOD USING GBFF AND STACKING AS THE FUSION STEP. THE NUMBER OF FEATURES IS WRITTEN IN PARENTHESES.

	LiDAR (1)	**Hyper** (144)	**LiDAR+Hyper** (145)	$\mathbf{EP}_{\text{lidar}}$ (71)	$\mathbf{EP}_{\text{hyper}}$ (213)	$\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ (284)	**GBFF** (50)	**Stack** (301)
OA	31.83	77.47	80.91	73.42	80.36	86.98	87.25	**88.91**
AA	37.43	80.34	83.17	75.97	83.47	88.54	88.95	**90.15**
Kappa	0.2677	0.7563	0.7931	0.712	0.7876	0.8592	0.8615	**0.8796**
1	13.48	83.38	**83.57**	74.26	77.49	78.06	80.06	83.29
2	16.25	**98.40**	98.21	61.75	78.48	84.96	92.58	97.74
3	56.63	98.02	98.42	97.23	**100.00**	**100.00**	**100.00**	**100.00**
4	44.03	97.54	97.73	58.14	82.77	95.45	95.55	**99.34**
5	58.04	96.40	96.50	82.10	97.73	98.77	**99.72**	98.77
6	58.04	97.20	97.20	83.22	95.80	95.80	95.80	**99.30**
7	39.08	82.09	85.82	77.33	73.23	73.41	**86.38**	85.91
8	29.53	40.65	56.51	68.28	59.92	85.28	86.61	**86.99**
9	13.59	69.78	71.20	59.40	83.00	**93.96**	91.31	91.97
10	11.29	57.63	57.14	66.89	64.09	**67.08**	47.49	49.71
11	40.41	76.09	80.55	**99.91**	84.72	90.89	92.88	97.82
12	9.99	49.38	62.82	64.75	78.10	**88.57**	85.11	86.26
13	15.08	61.40	63.86	58.60	77.89	76.14	**82.46**	75.44
14	80.16	99.60	**100.00**	**100.00**	99.60	**100.00**	99.60	**100.00**
15	80.16	97.67	98.10	87.74	99.37	**99.79**	98.73	**99.79**

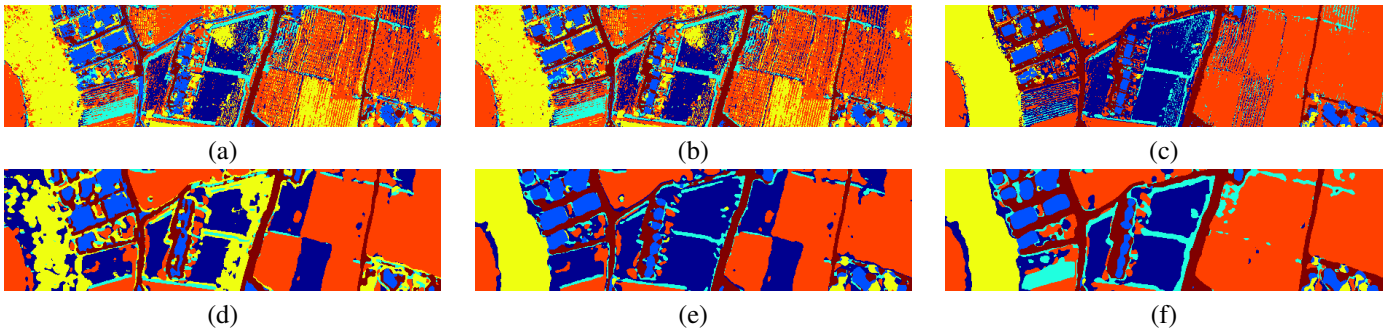

Fig. 9. Classification maps for Trento data: The outputs of RF on (a) hyperspectral data, (b) the stack of LiDAR and hyperspectral data, and (c) the proposed approach using feature stacking; the outputs of CNN-based classification on (d) hyperspectral data, (e) the stack of LiDAR and hyperspectral data, and (f) the proposed approach using GBFF.

feature extraction on the stacked features may improve the classification accuracy of CNN-based classifiers. With respect to using the CNN for the classification step, the graph-based feature fusion approach could lead to better results in terms of classification accuracies than feature stacking.

It should be noted that, in this work, deep learning has been used for the first time for the joint classification of LiDAR and hyperpspectral data in our community, and its results indicate that the convolutional neural network is an efficient tool for the fusion of LiDAR and hyperspectral data.

V. ACKNOWLEDGMENT

The authors would like to thank Prof. Lorezone Bruzzone of the University of Trento for providing the Trento data set. In addition, the authors would like to express their appreciation to the National Center for Airborne Laser Mapping (NCALM) for providing the Houston data set. The shadow-removed hyperspectral data is provided by Prof. Naoto Yokoya. Furthermore, the authors greatly appreciate Dr. Wenzhi Liao from Gent University for sharing the graph-based feature fusion code with us. This research has been partly supported by the Alexander von Humboldt Fellowship for postdoctoral researchers, and the Helmholtz Young Investigators Group "SiPEO" (VH-NG-1018, www.sipeo.bgu.tum.de).

REFERENCES

[1] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, W. Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Q. Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest," *IEEE Jour. Selec. Top. App. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, June 2014.

[2] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral-spatial classification of hyperspectral images

TABLE VI

CNN HOUSTON: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT APPROACHES USING CNN. THE METRICS AA AN OA ARE REPORTED IN PERCENTAGE. KAPPA COEFFICIENT IS OF NO UNITS. THE BEST RESULT IS SHOWN IN BOLD. **LIDAR**, **HYPER**, AND **LIDAR+HYPER** SHOW THE CLASSIFICATION ACCURACIES OF LIDAR, HYPERSPECTRAL, AND THEIR STACK, RESPECTIVELY. $\mathbf{EP}_{lidar}$, $\mathbf{EP}_{\text{hyper}}$, AND $\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ SHOW THE CLASSIFICATION ACCURACIES OF EPS APPLIED TO LIDAR, HYPERSPECTRAL, AND THE STACK OF MEP AND EMEP, RESPECTIVELY. **GBFF** AND **STACK** SHOW THE CLASSIFICATION ACCURACIES OF THE PROPOSED METHOD USING GBFF AND STACKING AS THE FUSION STEP. THE NUMBER OF FEATURES IS WRITTEN IN PARENTHESES.

	LiDAR (1)	**Hyper** (144)	**LiDAR+Hyper** (145)	$\mathbf{EP}_{\text{lidar}}$ (71)	$\mathbf{EP}_{\text{hyper}}$ (213)	$\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ (284)	**GBFF** (50)	**Stack** (301)
OA	49.79	78.35	83.33	61.76	88.01	88.81	**91.02**	89.71
AA	49.67	77.19	83.21	62.47	89.12	90.00	**91.82**	90.39
Kappa	0.4563	0.7646	0.8188	0.5851	0.8702	0.8788	**0.9033**	0.8884
1	28.30	82.24	**83.48**	52.52	78.16	78.63	78.73	78.35
2	26.88	**98.31**	89.10	38.06	87.41	94.83	94.92	94.64
3	49.50	70.69	83.17	71.49	99.80	99.80	**100.00**	**100.00**
4	62.03	94.98	**99.34**	73.86	79.64	98.30	**99.34**	99.15
5	28.60	97.25	97.63	29.92	98.39	98.58	**99.62**	98.77
6	37.76	79.02	**98.60**	66.43	94.41	95.10	95.80	95.10
7	52.71	86.19	**93.10**	87.03	79.76	82.56	87.87	85.45
8	77.30	65.81	88.03	82.72	94.11	92.02	**95.25**	92.88
9	49.34	72.11	76.47	51.23	81.31	85.39	**89.71**	83.78
10	64.48	55.21	43.92	57.82	**97.01**	71.04	81.18	81.76
11	71.35	85.01	91.46	88.71	86.15	85.77	86.34	84.91
12	43.32	60.23	75.70	54.95	91.55	92.12	**92.70**	92.03
13	38.95	75.09	74.74	67.02	**89.82**	87.02	87.02	86.32
14	87.04	83.00	82.19	73.68	95.14	98.38	**99.19**	94.33
15	27.48	52.64	71.25	41.65	84.14	**90.49**	89.64	88.37

TABLE VII

RF TRENTO: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT APPROACHES USING RF. THE METRICS AA AN OA ARE REPORTED IN PERCENTAGE. KAPPA COEFFICIENT IS OF NO UNITS. THE BEST RESULT IS SHOWN IN BOLD. **LIDAR**, **HYPER**, AND **LIDAR+HYPER** SHOW THE CLASSIFICATION ACCURACIES OF LIDAR, HYPERSPECTRAL, AND THEIR STACK, RESPECTIVELY. $\mathbf{EP}_{\text{lidar}}$, $\mathbf{EP}_{\text{hyper}}$, AND $\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ SHOW THE CLASSIFICATION ACCURACIES OF EPS APPLIED TO LIDAR, HYPERSPECTRAL, AND THE STACK OF MEP AND EMEP, RESPECTIVELY. **GBFF** AND **STACK** SHOW THE CLASSIFICATION ACCURACIES OF THE PROPOSED METHOD USING GBFF AND STACKING AS THE FUSION STEP. THE NUMBER OF FEATURES IS WRITTEN IN PARENTHESES.

	LiDAR (1)	**Hyper** (63)	**LiDAR+Hyper** (64)	$\mathbf{EP}_{\text{lidar}}$ (71)	$\mathbf{EP}_{\text{hyper}}$ (213)	$\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ (284)	**GBFF** (50)	**Stack** (290)
OA	46.7	84.92	90.61	95.9	85.17	98.39	97.66	**98.45**
AA	43.31	85.01	89.17	93.53	84.43	97.06	96.87	**97.17**
Kappa	0.335	0.8004	0.8566	94.53	0.8099	0.9785	0.9688	**0.9793**
1	42.5	86.2	86.09	97.82	96.06	97.62	**99.73**	98.19
2	51.3	85.9	93.87	94.25	98.42	96.80	**97.04**	96.56
3	34.2	96.8	**97.91**	94.99	72.03	94.36	95.82	94.78
4	52.6	95.7	97.05	99.22	99.45	**99.97**	**99.97**	99.92
5	46.5	80.1	82.76	98.76	69.89	99.10	96.90	**99.19**
6	32.4	65	86.01	76.15	70.79	**94.55**	91.78	94.42

TABLE VIII

CNN TRENTO: CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT APPROACHES USING CNN. THE METRICS AA AN OA ARE REPORTED IN PERCENTAGE. KAPPA COEFFICIENT IS OF NO UNITS. THE BEST RESULT IS SHOWN IN BOLD. **LIDAR**, **HYPER**, AND **LIDAR+HYPER** SHOW THE CLASSIFICATION ACCURACIES OF LIDAR, HYPERSPECTRAL, AND THEIR STACK, RESPECTIVELY. $\mathbf{EP}_{\text{lidar}}$, $\mathbf{EP}_{\text{hyper}}$, AND $\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ SHOW THE CLASSIFICATION ACCURACIES OF EPS APPLIED TO LIDAR, HYPERSPECTRAL, AND THE STACK OF MEP AND EMEP, RESPECTIVELY. **GBFF** AND **STACK** SHOW THE CLASSIFICATION ACCURACIES OF THE PROPOSED METHOD USING GBFF AND STACKING AS THE FUSION STEP. THE NUMBER OF FEATURES IS WRITTEN IN PARENTHESES.

	LiDAR (1)	**Hyper** (63)	**LiDAR+Hyper** (64)	$\mathbf{EP}_{lidar}$ (71)	$\mathbf{EP}_{\text{hyper}}$ (213)	$\mathbf{EP}_{\text{lidar}} + \mathbf{EP}_{\text{hyper}}$ (284)	**GBFF** (50)	**Stack** (290)
OA	69.93	83.52	97.48	95.88	90.72	98.70	**98.93**	98.85
AA	49.42	80.59	95.58	92.50	84.99	98.08	**98.48**	98.40
Kappa	0.5802	0.7843	0.9664	0.9450	0.8763	0.9827	**0.9855**	0.9846
1	7.54	92.22	95.88	**99.90**	98.02	99.63	99.67	99.53
2	59.87	87.08	99.07	99.14	98.79	**99.31**	98.53	98.79
3	4.80	66.81	91.44	90.40	70.77	99.37	**99.97**	99.79
4	91.13	65.24	99.79	**99.89**	99.41	99.80	99.72	99.50
5	89.03	98.98	98.56	99.02	90.65	99.60	99.52	**99.76**
6	44.17	73.19	88.72	66.67	52.30	90.74	**93.48**	93.01

based on hidden Markov random fields," *IEEE Trans. Geos. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, 2014.

[3] P. Ghamisi, M. S. Couceiro, F. M. L. Martins, and J. A. Benediktsson, "Multilevel image segmentation approach for remote sensing images based on fractional-order darwinian particle swarm optimization," *IEEE Trans. Geos. Remote Sens.*, vol. 52, no. 5, pp. 2382–2394, 2014.

[4] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geos. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, 2008.

[5] J. Li, J. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geos. Remote Sens.*, vol. 48, no. 11, p. 40854098, 2010.

[6] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. Geos. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, 2010.

[7] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Sys. Man. Cyber.*, 2010.

[8] S. Bernabé, P. R. Marpu, A. Plaza, M. Dalla Mura, and J. A. Benediktsson, "Spectral-spatial classification of multispectral images using kernel feature space representation," *IEEE Geos. Remote Sens. Let.*, no. 1, pp. 288 – 292, 2013.

[9] J. A. Benediktsson and P. Ghamisi, *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*. Artech House Publishers, INC, Boston, USA, 2015.

[10] M. Belgiu, I. Tomljenovic, T. J. Lampoltshammer, T. Blaschke, and B. Höfle, "Ontology-based classification of building types detected from airborne laser scanning data," *Remote Sens.*, vol. 6, no. 2, pp. 1347–1366, 2014.

[11] A. Jochem, B. Höfle, M. Rutzinger, and N. Pfeifer, "Automatic roof plane detection and analysis in airborne lidar point clouds for solar potential assessment," *Sensors*, vol. 9, no. 7, pp. 5241–5262, 2009.

[12] M. Rutzinger, B. Höfle, M. Hollaus, and N. Pfeifer, "Object-based point cloud analysis of full-waveform airborne laser scanning data for urban vegetation classification," *Int. Jour. Image Data Fus.*, vol. 8, no. 8, pp. 4505–4528, 2008.

[13] R. K. Hall, R. L. Watkins, D. T. Heggem, K. B. Jones, P. R. Kaufmann, S. B. Moore, and S. J. Gregory, "Quantifying structural physical habitat attributes using lidar and hyperspectral imagery," *Int. Jour. Image Data Fus.*, vol. 59, no. 1, pp. 63–83, 2009.

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

[15] I. Tomljenovic, B. Höfle, D. Tiede, and T. Blaschke, "Building extraction from airborne laser scanning data: An analysis of the state of the art," *Remote Sens.*, vol. 7, no. 4, pp. 3826–3862, 2015.

[16] B. Höfle, M. Hollaus, and J. Hagenauer, "Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne lidar data," *ISPRS Jour. Photo. Remote Sens.*, vol. 67, pp. 134–147, 2012.

[17] P. Gamba, F. Dell'Acqua, and B. V. Dasarathy, "Urban remote sensing using multiple data sets: Past, present, and future," *IEEE Trans. Geos. Remote Sens.*, vol. 6, p. 319326, 2005.

[18] Q. Chen, "Airborne lidar data processing and information extraction," *Photo. Eng. & Remote Sens.*, vol. 73, no. 2, pp. 109 – 112, 2007.

[19] P. Ghamisi, J. A. Benediktsson, and S. Phinn, "Land-cover classification using both hyperspectral and lidar data," *Int. Jour. Image Data Fus.*, vol. 6, no. 3, pp. 189–215, 2015.

[20] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT, no. 14, pp. 55 – 63, 1968.

[21] P. Ghamisi, "Spectral and spatial classification of hyperspectral data," Ph.D. dissertation, University of Iceland, 2015.

[22] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote-sensing data over urban areas," *EURASIP Jour. Adv. Sig. Proc.*, 2009.

[23] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geos. Remote Sens. Let.*, vol. 12, no. 2, pp. 309 – 313, 2015.

[24] P. Ghamisi, M. S. Couceiro, and J. A. Benediktsson, "A novel feature selection approach based on FODPSO and SVM," *IEEE Trans. Geos. Remote Sens.*, vol. 53, no. 5, pp. 2935–2947, 2015.

[25] Y. Bazi and F. Melgani, "Toward an optimal svm classification system for hyperspectral remote sensing images," *IEEE Trans. Geos. Remote Sens.*, vol. 44, no. 11, pp. 3374–3385, 2006.

[26] G. P. Asner, D. E. Knapp, T. Kennedy-Bowdoin, M. O. Jones, R. E. Martin, and J. Boardman, "Invasive species detection in hawaiian rainforests using airborne imaging spectroscopy and lidar," *Remote Sens. Env.*, vol. 112, pp. 1942–1955, 2008.

[27] G. A. Blackburn, "Remote sensing of forest pigments using airborne imaging spectrometer and lidar imagery," *Remote Sens. Env.*, vol. 82, p. 311321, 2002.

[28] M. Voss and R. Sugumaran, "Seasonal effect on tree species classification in an urban environment using hyperspectral data, lidar, and an object-oriented approach," *Sensors*, vol. 8, pp. 3020–3036, 2008.

[29] R. M. Lucas and A. C. L. andP. J. Bunting, "Retrieving forest biomass through integration of casi and lidar data," *Int. Jour. Remote Sens.*, vol. 29, pp. 1553–1577, 2008.

[30] U. Heiden, W. Heldens, S. Roessner, K. Segl, T. Esch, and A. Mueller, "Urban structure type characterization using hyperspectral remote sensing and height information," *Landsc. Urban Plann.*, vol. 105, no. 6, pp. 361–375, June 2012.

[31] R. K. Hall, R. L. Watkins, D. T. Heggem, K. B. Jones, P. R. Kaufmann, and S. B. Moore, "Quantifying structural physical habitat attributes using lidar and hyperspectral imagery," *Env. Mon. Asses.*, vol. 159, pp. 63–83, 2009.

[32] B. Koetz, F. Morsdorf, S. Linder, T. Curt, and B. Allgower, "Multi-source land cover classification for forest fire management based on imaging spectrometry and lidar data." *For. Ecol. Management*, vol. 256, pp. 263–271, 2008.

[33] J. T. Mundt, D. R. Streutker, and N. F. Glenn, "Mapping sagebrush distribution using fusion of hyperspectral and lidar classifications," *Phot. Eng. Remote Sens.*, vol. 72, no. 1, p. 4754, Sept 2006.

[34] 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*, p. 16, April 2007.

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

[36] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geos. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, 2010.

[37] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles,," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, p. 480491, 2005.

[38] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Trans. Geos. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, 2015.

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

[40] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geos. Remote Sens. Let.*, vol. 13, no. 11, pp. 1641–1645, 2016.

[41] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Comp.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[42] R. Salakhutdinov and G. E. Hinton, "Deep boltzmann machines," *Int. Conf. Art. Intel. Stat.*, pp. 448–455, Clearwater Beach, Florida, 2009.

[43] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *Neural Inf. Proc. Sys. 19*, pp. 153–160, Cambridge, USA, 2007.

[44] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2094 – 2107, 2014.

[45] Y. Chen, X. Zhao, and X. Jia, "Spectra-spatial classification of hyperspectral data based on deep belief network," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2381–2292, 2015.

[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Proc. Adv. Neural Inf. Process*, pp. 1097 –1105, 2012.

[47] P. Ghamisi, Y. Chen, and X. X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geos. Remote Sens. Let.*, vol. 13, no. 10, pp. 1537–1541, 2016.

[48] W. Liao, R. Bellens, S. Gautama, and W. Philips, "Feature fusion of hyperspectral and lidar data for classification of remote sensing data from urban area," in *EARSeL Special Interest Group on Land Use and Land Cover, 5th Workshop*, S. V. D. Linden, T. Kuemmerle, and K. Janson, Eds., 2014, pp. 34–34.

[49] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, p. 532, 2001.

[50] C. Vachier, "Extinction value: a new measurement of persistence," in *IEEE Workshop on Nonlinear Signal and Image Processing*, vol. I, 1995, pp. 254–257.

[51] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.

[52] G. Matheron, "Elements pour une theprie des milieux poreus," *Mason*, 1967.

[53] R. Souza, L. Rittner, R. Machado, and R. Lotufo, "A comparison between extinction filters and attribute filters," in *ISMM'15*, 2015, pp. 63–74.

[54] E. Carlinet and T. Geraud, "A comparative review of component tree computation algorithms," *IEEE Trans. Image Proc.*, vol. 23, no. 9, pp. 3885–3895, Sept 2014.

[55] R. Souza, L. Rittner, R. Lotufo, and R. Machado, "An array-based node-oriented max-tree representation," in *ICIP'15*, Sept 2015, pp. 3620–3624.

[56] Y. LeCun and Y. Bengio, "The handbook of brain theory and neural networks," M. A. Arbib, Ed. Cambridge, MA, USA: MIT Press, 1998, ch. Convolutional Networks for Images, Speech, and Time Series, pp. 255–258. [Online]. Available: http://dl.acm.org/citation.cfm?id=303568.303704

[57] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *ICANN*, 2010.

[58] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geos. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016.

[59] P. Ghamisi, J. A. Benediktsson, G. Cavallaro, and A. Plaza, "Automatic framework for spectral-spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 7, no. 6, pp. 2147–2160, June 2014.

[60] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction,"

IEEE Trans. Geos. Remote Sens., vol. 52, no. 5, pp. 342–346, 2014.

[61] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comp.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[62] M. Pedergnana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and lidar data using extended attribute profiles," *IEEE Jour. Sel. Top. Signal Proc.*, vol. 6, no. 7, pp. 856–865, 2012.

[63] M. Khodadadzadeh, J. Li, S. Prasad, and A. Plaza, "Fusion of hyperspectral and lidar remote sensing data using multiple feature learning," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2971–2983, 2015.

Pedram Ghamisi (S'13, M'15) graduated with a B.E. in civil (survey) engineering from the Tehran South Campus of Azad University. He obtained an M.E. degree with first class honors in remote sensing at K.N.Toosi University of Technology in 2012. In 2013/2014, he spent seven months at the school of Geography, Planning and Environmental Management, the University of Queensland, Australia. He received a Ph.D. in electrical and computer engineering at the University of Iceland, Reykjavik in 2015 and subsequently worked as a postdoctoral research fellow at the University of Iceland. In 2015, Dr. Ghamisi won the prestigious Alexander von Humboldt Fellowship and started his work as a postdoctoral research fellow at Technical University of Munich (TUM) and Heidelberg University, Germany from October 2015. He has also been working as a researcher at German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Germany since October 2015. His research interests are in remote sensing and image analysis, with a special focus on spectral and spatial techniques for hyperspectral image classification and the integration of LiDAR and hyperspectral data for land cover assessment.

In the academic year 2010-2011, he received the Best Researcher Award for M.Sc. students in K. N. Toosi University of Technology. At the 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, July 2013, Dr. Ghamisi was awarded the IEEE Mikio Takagi Prize for winning the Student Paper Competition, competing with almost 70 submissions. In 2016, he was selected as *talented international researcher* by Iran's National Elites Foundation.

Bernhard Höfle received his PhD degree from the University of Innsbruck, Austria, in 2007. He is currently professor of GIScience and 3-D spatial data processing at the Institute of Geography, Heidelberg University, Germany. He also is member of the executive board of the Heidelberg Center for the Environment, co-chair of the working group geoinformatics of the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF), and council member of the International Sociecty for Digital Earth (ISDE). His research interests comprise GIS algorithms, object-based image and 3-D point cloud analysis, multisource geoinformation fusion, 3-D SDI, LiDAR applications in geosciences, and radiometric calibration and analysis of LiDAR data.

Xiaoxiang Zhu (S'10-M'12-SM'14) received the bachelor degree in space engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2006. She received the Master (M.Sc.) degree, her doctor of engineering (Dr.-Ing.) degree and her Habilitation in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011 and 2013, respectively. Since 2011, she is a scientist with the Remote Sensing Technology Institute at the German Aerospace Center (DLR), Oberpfaffenhofen, where she is the head of the Team Signal Analysis. Since 2013, she is also a Helmholtz Young Investigator Group Leader and appointed as TUM junior fellow. In 2015, she is appointed as the Professor for Signal Processing in Earth Observation at TUM. Prof. Zhu was a guest scientist or visiting professor at the Italian National Research Council (CNR-IREA), Naples, Italy, Fudan University, Shanghai, China and the University of Tokyo, Tokyo, Japan in 2009, 2014 and 2015, respectively.

Her main research interests are: advanced InSAR techniques such as high dimensional tomographic SAR imaging and SqueeSAR; computer vision in remote sensing including object reconstruction and multi-dimensional data visualization; and modern signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing such as multi/hyperspectral image analysis.

Dr. Zhu is an associate Editor of IEEE Transactions on Geoscience and Remote Sensing.