Fusion of Hyperspectral and LiDAR Data for Classification of Cloud-Shadow Mixed Remote Sensed Scene

Renbo Luo, *Student Member, IEEE*, Wenzhi Liao, *Senior Member, IEEE*, Hongyan Zhang, *Senior Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, Paul Scheunders, *Senior Member, IEEE*, Youguo Pi, and Wilfried Philips, *Senior Member, IEEE*

***Abstract*—Recent advances in sensor design allow us to gather more useful information about the Earth's surface. Examples are hyperspectral (HS) and Light Detection And Ranging (LiDAR) sensors. These, however, have limitations. HS data cannot distinguish different objects made from similar materials and highly suffers from cloud-shadow regions, whereas LiDAR cannot separate distinct objects that are at the same altitude. For an increased classification performance, fusion of HS and LiDAR data recently attracted interest but remains challenging. In particular, these methods suffer from a poor performance in cloud-shadow regions because of the lack of correspondence with shadow-free regions and insufficient training data. In this paper, we propose a new framework to fuse HS and LiDAR data for the classification of remote sensing scenes mixed with cloud-shadow. We process the cloud-shadow and shadow-free regions separately, our main contribution is the development of a novel method to generate reliable training samples in the cloud-shadow regions. Classification is performed separately in the shadow-free (classifier is trained by the available training samples) and cloud-shadow regions (classifier is trained by our generated training samples) by integrating spectral (i.e., original HS image), spatial (morphological features computed on HS image) and elevation (morphological features computed on LiDAR) features. The final classification map is obtained by fusing the results of the shadow-free and cloud-shadow regions. Experimental results on a real HS and LiDAR dataset demonstrate the effectiveness of the proposed method, as the proposed framework improves the overall classification accuracy with 4% for whole scene and 10% for shadow-free regions over the other methods.**

***Index Terms*—Classification, cloud–shadow, feature fusion, hyperspectral (HS) image, Light Detection And Ranging (LiDAR).**

Manuscript received November 18, 2016; revised January 20, 2017 and March 6, 2017; accepted March 13, 2017. This work was supported in part by FWO under Project G037115N: Data fusion for image analysis in remote sensing, in part by the China Scholarship Council, and in part by the Open Fund of State Key Laboratory of Remote Sensing Science under Grant OFSLRSS201610. The work of W. Liao was supported by the Research Foundation Flanders (FWO-Vlaanderen). *(Corresponding author: Renbo Luo.)*

R. Luo is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China, and also with the Department of Telecommunications and Information Processing, Ghent University, Ghent 9000, Belgium (e-mail: renbo.luo@ugent.be).

W. Liao is with the Department of Telecommunications and Information Processing, Ghent University, Ghent 9000, Belgium, and also with the State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China (e-mail: wenzhi.liao@ugent.be).

H. Zhang and L. Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zhanghongyan@whu.edu.cn; zlp62@whu.edu.cn).

P. Scheunders is with Vision Laboratory, University of Antwerp, Antwerp 2610 Belgium (e-mail: paul.scheunders@uantwerpen.be).

Y. Pi is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: auygpi@scut.edu.cn).

W. Philips is with the Department of Telecommunications and Information Processing, Ghent University, Ghent 9000, Belgium (e-mail: wilfried.philips@ugent.be).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2017.2684085

I. Introduction

RECENT advances in sensor technology allow us to measure different aspects of the objects on the Earth's surface, e.g., the spectral radiance using hyperspectral (HS) images, and height information using Light Detection And Ranging (LiDAR) data [1]. Nowadays, HS images of both high spatial and spectral resolutions are available and can provide valuable spectral information for land use/cover applications [2]. However, their use is still limited in very complex scenes in which many objects are made up of similar materials (e.g., roofs, parking lots, and roads) [3], [4]. Moreover, optical in nature, HS sensors suffer from cloudy weather conditions. On the other hand, LiDAR image provides complementary information related to the size, structure, and elevation of different objects [5], but fails to discriminate between different objects that are similar in altitude while quite different in nature (e.g., grass field and swimming pool). Therefore, using a single data source (either HS or LiDAR data) alone might not be sufficient to obtain reliable classification results.

Due to an increased availability of HS and LiDAR data from overlapping areas, the fusion of HS and LiDAR data has recently been explored intensively. In [6], Gu *et al.* proposed a multiple-kernel learning model to integrate heterogeneous features from HS and LiDAR data for urban area classification. Elakshe *et al.* [7] explored the fusion of HS and LiDAR data for coastal mapping by using HS imagery to discriminate between road and water pixels, and LiDAR image to detect and create a vector layer of building polygons. Dalponte *et al.* [8] investigated the joint use of HS and LiDAR data for the classification of complex forest areas. Yokoya *et al.* [9] fused HS

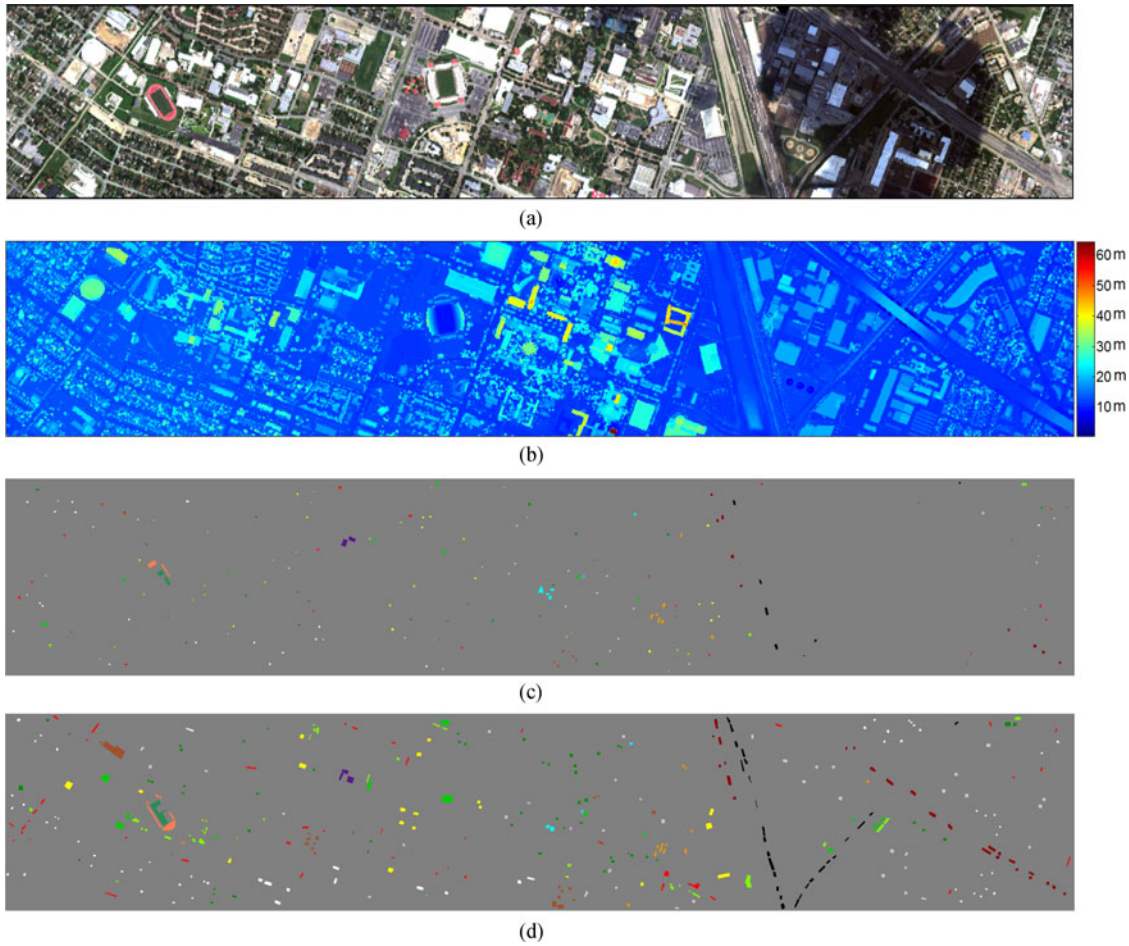


Fig. 1. (a) False RGB image of HS data; (b) LiDAR image; (c) available training samples; and (d) testing samples.

and LiDAR data for landscape visual quality assessment and enabled the prediction of landscape quality from any viewpoint using large-scale remote sensing observations. In [10], classification of eight common savanna tree species was performed by fusing HS and LiDAR data with an automated random forest modeling approach. Shimoni *et al.* proposed a score-level fusion approach to detect stationary vehicles under shadows in [11], where detection scores from both HS and LiDAR data are derived separately and combined with a simple sum rule. It is clear that the combination of HS and LiDAR data can contribute to a more comprehensive interpretation of ground objects [6]–[13].

As the footprint of one object often contains more than one pixel and thus a high spatial correlation is expected between neighboring pixels, many approaches [14]–[16] incorporated spatial information to improve the fusion process of HS and LiDAR data and address the salt and pepper phenomenon in classification. In [12], Pedergnana *et al.* applied morphological attribute profiles (MAPs) [17] to model information from both HS and LiDAR data, and fused multiple feature sources in a stacked architecture. Recently, Khodadadzadeh *et al.* [18] developed a new strategy to fuse HS and LiDAR data by stacking multiple types of features (spatial and spectral features from HS, elevation features from LiDAR). The methods mentioned above have demonstrated that combining spectral, spatial, and elevation features further boosts the accuracy of land cover classification maps. However, stacking the high-dimensional spectral and morphological features directly may lead to the curse of dimensionality problem [19] and excessive computation time.

In 2013, the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) organized a contest involving two types of data sources: a cloud-shadow HS image and a LiDAR-derived digital surface model (DSM) [20] (see Fig. 1). The competition was established to stimulate the development of advanced methods to fuse HS and LiDAR data for classification [21]. More than 900 researchers from universities, national labs, space agencies, and corporations across the globe registered to the contest and provided solutions. In particular, a graph-based fusion method [21] and its generalized version [22] were proposed to couple dimensionality reduction and feature fusion of the spectral information (of the original HS image) and MAPs (built on both HS and LiDAR data). Debes *et al.* [21] proposed a two-stream classification framework which combined the HS and LiDAR data by a parallel process that involves both unsupervised and supervised classification. In very recent work, Ghamisi *et al.* [25] fused HS and LiDAR data by using extinction profiles and deep convolutional neural networks and achieved improved classification results.

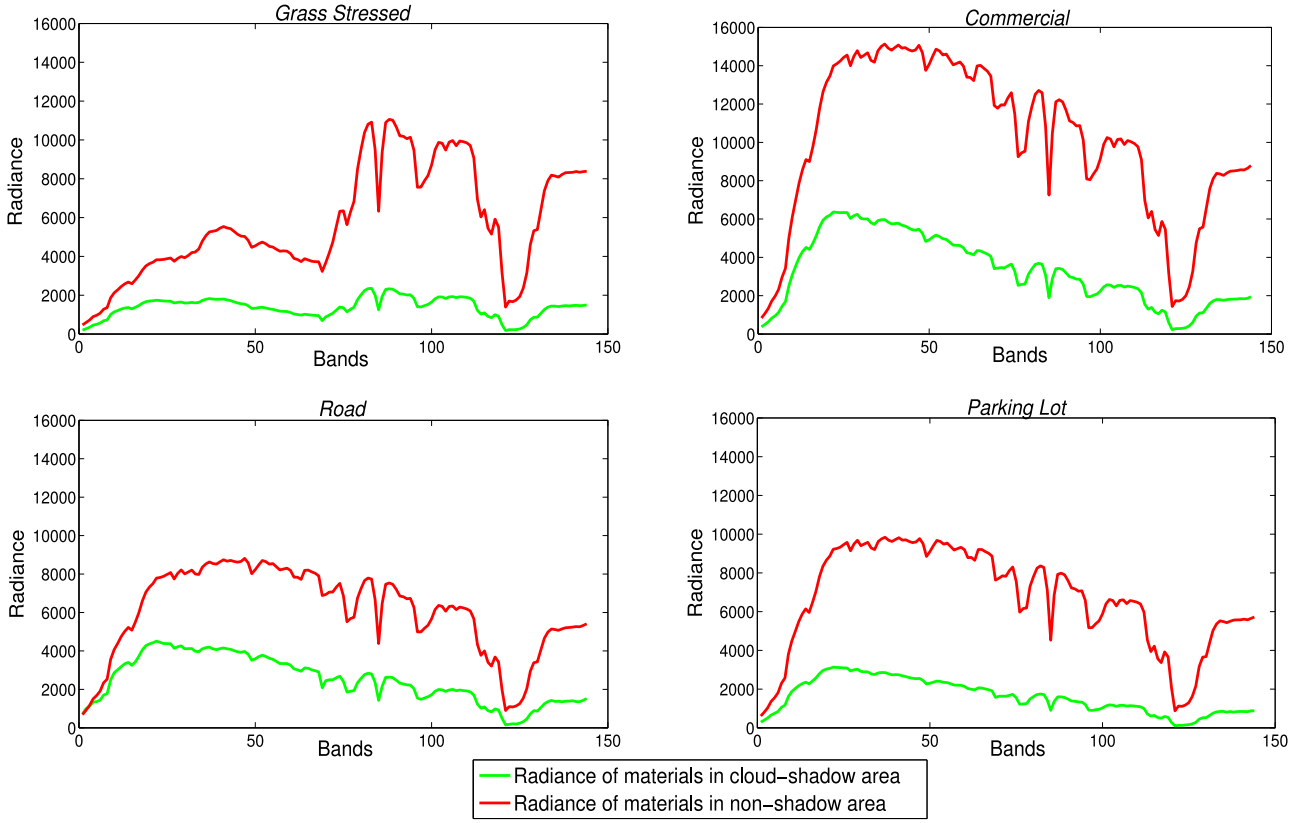


Fig. 2. Radiance (at-sensor spectral radiance units: $\mu\mathrm{W}\cdot\mathrm{cm}^{-2}\cdot\mathrm{sr}^{-1}\cdot\mathrm{nm}^{-1}$) of different materials.

It is noteworthy that a large cloud shadow was present during the acquisition of the HS image. As cloud shadows blur most of the spectral information, they prevent accurate land cover mapping [26]. Moreover, as the clouds emerge and move irregularly and unpredictably, it is very difficult to label training samples and acquire remote sensing images at the same time, and to prepare two distinct sets of training samples for shadow-free and cloud-shadow regions in a remote sensing scene. Typically, most of the pixels in the cloud-shadow regions will be misclassified when only using training samples selected from shadow-free regions, as the spectral radiance information of samples located within and out of the cloud-shadow are totally different (see Fig. 2). For example, even though all the available training samples located in shadow-free regions were used to train the classifiers discussed in [18]–[24], the classification performances of the cloud-shadow regions were not satisfactory.

In this paper, we propose a novel framework to fuse HS and LiDAR data for the classification of remote sensing scenes mixed with cloud shadow. The proposed method performs classification separately on the cloud-shadow and shadow-free regions. We solve the problem of missing training samples in the cloud-shadow regions by developing a method to generate reliable training samples. This method is based on the assumption that in the cloud-shadow regions different feature sources share similar intracluster distance relations, and that elevation features are more reliable than spectral and/or spatial features. We then classify the shadow-free (using the available training samples) and cloud-shadow regions (using the generated training samples) separately by integrating spectral, spatial, and elevation features, obtained by exploiting attribute profiles (AP) [27]. The final classification map is produced by the decision fusion of the obtained cloud-shadow and shadow-free maps.

We can also interpret our proposed framework from the viewpoint of domain adaptation [28]. The shadow-free region can be seen as the source domain, whereas the cloud-shadow region can be seen as a target domain. The labeled training set is only available for the source domain. According to the condition that the source and target domains share the same set of classes and elevation features (LiDAR), we make use of the information from the source domain to select training samples for the classification of the target domain.

The remainder of this paper is organized as follows: Section II describes the proposed framework, with a detailed description of every part of the proposed method. The experimental results on real urban cloud-shadow HS and LiDAR data are presented and discussed in Section III. Finally, the conclusions of the paper are drawn in Section IV.

II. Proposed Framework

Clouds heavily distort the Sun's reflectance and information analysis based on such distorted optical images is not always reliable. Meanwhile, collection of training data is preferably done on the ground, and since the clouds emerge and move irregularly and unpredictably, it is very difficult to obtain remote sensing images and label training samples from both cloud-shadow and

TABLE I
SOME NOTATIONS USED IN THIS PAPER

Notations	Description
HS image	Raw data cube: $M \times N \times D$, D is the number of the spectral bands
$\text{EMAPs}_{\text{hsi}}$	AP generated from HS image: $M \times N \times D_1$, D_1 is the dimension of $\text{EMAPs}_{\text{hsi}}$
MAPs_{lid}	AP generated from LiDAR image: $M \times N \times D_2$, D_2 is the dimension of MAPs_{lid}
n	Number of labeled training samples
C	Number of classes
y_i	Label of ith training sample
$y_i^{'}$	Label of ith sample in Map_{lid}
$\mathbf{x}_i^{\text{spe}}$	ith sample in HS image, $\mathbf{x}_i^{\text{spe}} \in \mathbb{R}^D$
$\mathbf{x}_i^{\text{spa}}$	ith sample in $\text{EMAPs}_{\text{hsi}}$, $\mathbf{x}_i^{\text{spa}} \in \mathbb{R}^{D_1}$
$\mathbf{x}_i^{\text{Sta}}$	ith sample in the profile by stacking HS image and $\text{EMAPs}_{\text{hsi}}$, $\mathbf{x}_i^{\text{Sta}} = \{\mathbf{x}_i^{\text{spe}}; \mathbf{x}_i^{\text{spa}}\} \in \mathbb{R}^{D+D_1}$
d	The number of extracted features from each data source
$F\text{spe}$	Spectral features extracted from HS image: $M \times N \times d$
$F\text{spa}$	Spatial features extracted from $\text{EMAPs}_{\text{hsi}}$: $M \times N \times d$
$F\text{lid}$	Elevation features extracted from MAPs_{lid}: $M \times N \times d$
$\mathbf{G} = \{g_{ij}\}$	Cloud-shadow mask
$\mathbf{X}_{c(k)}^{'}$	Cotraining samples for class c in kth iteration
n_c^k	Number of samples in $\mathbf{X}_{c(k)}^{'}$
$\mathbf{m}_{c(k)}^{\text{spe}}$	Center of $\mathbf{X}_{c(k-1)}^{'}$ in spectral feature space, $\mathbf{m}_{c(k)}^{\text{spe}} \in \mathbb{R}^D$
$\mathbf{m}_{c(k)}^{\text{spa}}$	Center of $\mathbf{X}_{c(k-1)}^{'}$ in spatial feature space, $\mathbf{m}_{c(k)}^{\text{spa}} \in \mathbb{R}^{D_1}$

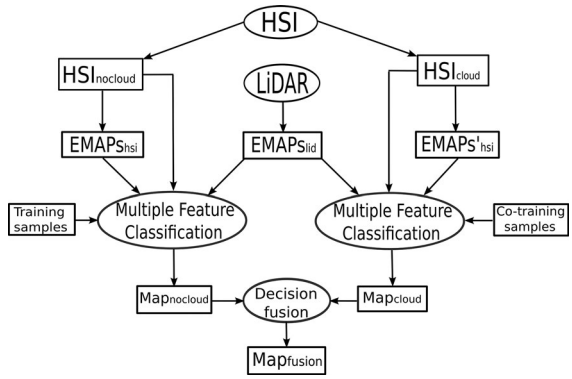


Fig. 3. Flowchart of the proposed framework, here $\text{HSI}_{\text{nocloud}}$ and $\text{HSI}_{\text{cloud}}$ denote the shadow-free and cloud-shadow regions of HS image, EMAP_{hsi} and EMAP_{hsi} mean MPs extracted from hyperspecral and LiDAR data. "Cotraining samples" are generated training samples by our methods for cloud-shadow regions.

shadow-free regions at the same time. Moreover, users prefer to select training samples from shadow-free regions for better visualization and interpretation. For example in Fig. 1(c), all training samples were collected from shadow-free regions. When classifying a remote sensing scene with a classifier, trained on a shadow-free training set only, the results on the cloud-shadow regions will be very poor, the main reason being that objects made up of the same material have different spectral signatures in cloud-shadow and shadow-free regions (see Fig. 2). However, it is important to notice that within the cloud-shadow regions, objects made up of different materials have different spectral signatures (see Fig. 2), indicating that these regions still contain sufficient distinctive information. Some notations used throughout this paper are summarized in Table I.

Therefore, we propose a novel framework for the classification of remote sensing scenes containing cloud shadows. In the proposed framework, as shown in Fig. 3, we first divide the HS image into two different parts: cloud-shadow ($\text{HSI}_{\text{cloud}}$) and shadow-free ($\text{HSI}_{\text{nocloud}}$) regions. $\text{EMAPs}_{\text{hsi}}$ and MAPs_{lid} denote the additional spatial and elevation information extracted from HS and LiDAR data by AP [27], respectively. For the classification of shadow-free regions $\text{HSI}_{\text{nocloud}}$, we fuse multiple features using a similar framework as in [29]. To reduce the information redundancy, we first use feature extraction (FE) techniques to extract relevant information from each single feature source. Then, we concatenate all extracted features together, and use these as input for a classifier to obtain the classification map of shadow-free regions $\text{Map}_{\text{nocloud}}$. In order to generate a classification map of cloud-shadow regions $\text{Map}_{\text{cloud}}$, we propose a novel method to select new training samples from cloud-shadow regions, which will be detailed in the following section. Using these training samples, the same multiple feature classification method as with the shadow-free regions is applied. Last but not least, the final classification map of a cloud mixed remote sensing scene is obtained by fusing $\text{Map}_{\text{nocloud}}$ and $\text{Map}_{\text{cloud}}$.

A. Morphological Attribute Profiles (MoAP)

For the classification of very high-resolution remote sensing data, spatial information has been widely exploited [30]–[33]. To model the spatial information from HS and LiDAR data, Pesaresi *et al.* [30] build so-called morphological profiles (MP). As an extension of the concept of MP, AP [27] provide a multilevel characterization of an image by the sequential application of morphological attribute filters, which model different specifications of the structural information contained in the scene, such as length, area, and shape of objects. In [31] and [32], extended multiattribute profiles (EMAPs) were developed to extract abundant spatial information in HS images and these profiles make an obvious contribution to the classification.

AP generate a multilevel decomposition of the input image based on attribute filters, which can properly extract and model the spatial information of the adjacent pixels with progressively higher threshold values. Suppose $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ ($\lambda_i <$

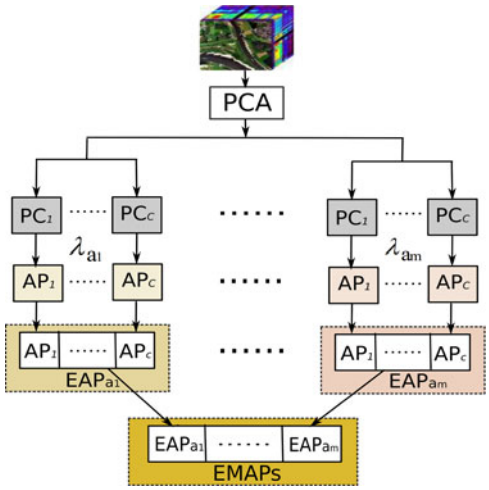


Fig. 4. General architecture of EMAPs, λ_{a_i} denotes the predefined conditions of attribute a_i.

λ_j with $i < j$) is a sequence of ordered predefined criteria, for a gray scale image g, then an AP of g can be defined as

$$\mathbf{AP}(g) = \{\phi_n(g), \ldots, \phi_1(g), g, \varphi_1(g), \ldots, \varphi_n(g)\} \quad (1)$$

where ϕ_i and φ_i denote the attribute thinning and thickening operations with reference values λ_i, respectively.

The above AP only works on a gray scale image. In order to extend the concept of the AP to HS image, one possible way is to perform a feature reduction approach (such as PCA) on the input data and then apply APs to the first principle components [32]. Let PC_i $i = \{1, \ldots, c\}$ denotes the first principle components of HS image, then extended-AP (EAP) can be mathematically given as

$$\mathbf{EAP} = \{\mathbf{AP}(\mathrm{PC}_1), \mathbf{AP}(\mathrm{PC}_2), \ldots, \mathbf{AP}(\mathrm{PC}_c)\}. \quad (2)$$

The presented EAPs model the size and structure of different objects based on one attribute. If more attributes (e.g., area, diagonal of bounding box, length, and standard deviation) are considered, EMAPs can be denoted as

$$\mathbf{EMAPs}_{\mathrm{hsi}} = \{\mathbf{EAP}_{a_1}, \overline{\mathbf{EAP}}_{a_2}, \ldots, \overline{\mathbf{EAP}}_{a_m}\} \quad (3)$$

where a_i is a generic attribute and $\overline{\mathbf{EAP}} = \mathbf{EAP} \setminus (\mathrm{PC}_1, \mathrm{PC}_2, \ldots, \mathrm{PC}_c)$, deleting PC$s$ from $\mathrm{EAP}_{a_i}, i > 1$ is necessary for avoiding redundancy since the original components PC_i are present in each EAP. Fig. 4 shows the general architecture of $\mathrm{EMAPs}_{\mathrm{hsi}}$.

AP can also be applied to model elevation information from LiDAR image. An attribute thinning acts on bright objects (for LiDAR image, the bright regions are actually areas with high elevation, such as the top of a roof), while thickening act on dark (low height) objects. For example, an attribute thinning deletes bright objects that are smaller than the threshold λ_i. By computing a series of attributes, a complete AP is built, carrying information about the elevation information of objects in the image. Suppose L denotes a LiDAR image, L can be seen as a gray scale image where the value of a pixel denotes the altitude at that point. Then, the AP of L can be defined as

$$\mathbf{AP}(L) = \{\phi_n(L), \ldots, \phi_1(L), L, \varphi_1(L), \ldots, \varphi_n(L)\}. \quad (4)$$

As LiDAR-derived image has only single band, we use the term multi-attribute profiles (MAPs) when different types of APs are applied to the LiDAR image. Then, the MAPs from LiDAR image ($\mathbf{MAPs}_{\mathrm{lid}}$) can be expressed as

$$\mathbf{MAPs}_{\mathrm{lid}} = \{\mathbf{AP}_{a_1}(L), \mathbf{AP}_{a_2}(L), \ldots, \mathbf{AP}_{a_m}(L)\} \quad (5)$$

where $a_i, i = \{1, \ldots, m\}$ denotes different types of attributes.

Fig. 5 shows some of the obtained APs on the HS and LiDAR image. The objects of the HS image under cloud shadow appear to be darker. Moreover, many objects (even in different categories) exhibit similar intensities [see Fig. 5(a)]. APs of LiDAR image are clearly less influenced by the cloud [see Fig. 5(b)], small objects disappear as the scale increases. With EMAP, additional spatial and elevation can be extracted.

B. Multiple Features Classification

From HS and LiDAR data, three types of feature sources can be obtained: spectral values from the original spectrum of the HS image, $\mathrm{EMAPs}_{\mathrm{hsi}}$ from the HS image, and $\mathrm{MAPs}_{\mathrm{lid}}$ from the LiDAR image, all of them having high dimensionality. If these features would be fused using a stacked structure, the dimensionality of this stacked vector will be very large, thus leading to the problem of the curse of dimensionality. Moreover, the stacked vector will contain redundancy and noisy. Therefore, it is necessary to use FE methods to reduce the dimensionality of the spectral features, the $\mathrm{EMAPs}_{\mathrm{hsi}}$ and $\mathrm{MAPs}_{\mathrm{lid}}$ before fusing the extracted low-dimensional features together for classification. Here, nonparametric weighted feature extraction (NWFE) [34] is chosen, as it is proven to be efficient to extract discriminative features for the classification of HS image [32].

Fig. 6(a) shows the proposed multiple feature classification strategy. First, the original HS data are transformed by FE to obtain a reduced set of effective spectral features (Fspe) that contains the spectral information of the HS data. In parallel, the HS image is transformed by PCA, and the first few important PCs that correspond to 99% of the cumulative variance are used to construct the $\mathrm{EMAPs}_{\mathrm{hsi}}$. If there are c PCs, each AP is composed of n thickening and n thinning transformations of the corresponding PC for each attribute and the number of attributes is m, then there are in total $c \times (m \times (2n) + 1)$ features in the EMAPs. In order to reduce redundancy and noise, avoid the curse of dimensionality and save processing time, FE is applied to extract an effective feature set (Fspa) from $\mathrm{EMAPs}_{\mathrm{hsi}}$ before classification. On the LiDAR image, exactly the same is done to extract an effective elevation feature set (Flid) from the $\mathrm{MAPs}_{\mathrm{lid}}$. Finally, the obtained Fspe, Fspa, and Flid are concatenated into one stacked vector Ffusion.

Another strategy would be to make up a large stacked vector from the spectral features, the spatial profiles of $\mathrm{EMAPs}_{\mathrm{hsi}}$ and elevation profiles of $\mathrm{MAPs}_{\mathrm{lid}}$ [18], and then extract effective features from this large stacked vector [see Fig. 6(b)]. In that case, however, since the different feature sources have

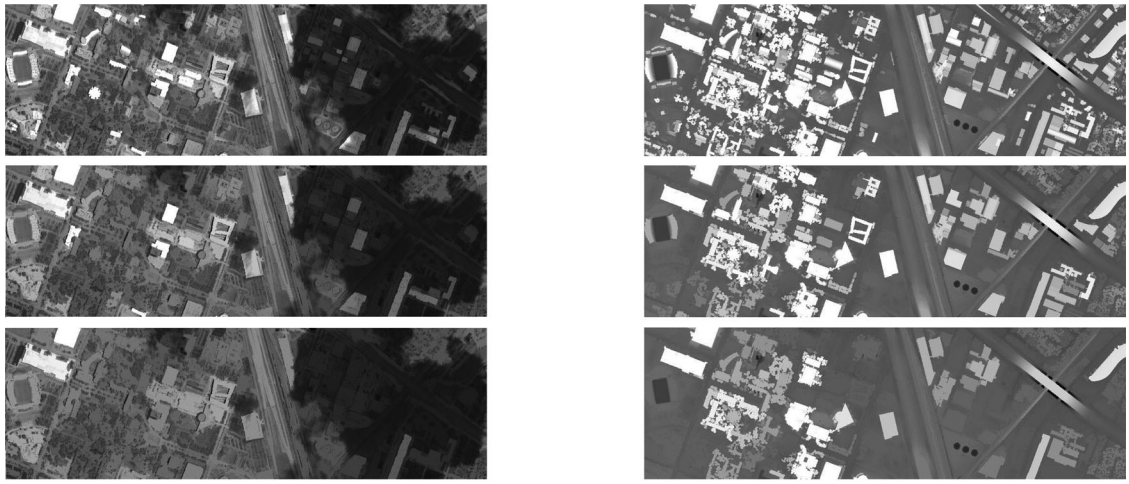


Fig. 5. Attribute thinning with "area" attribute. From up to down, the area size was set to 200, 500, and 1000, respectively. (a) APs of HS image with first PC and (b) APs of LiDAR image.

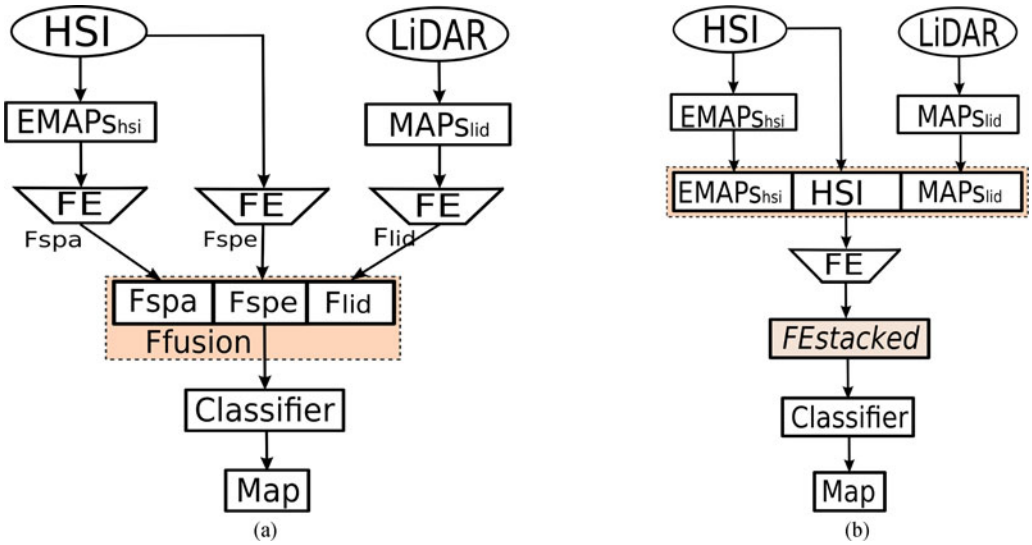


Fig. 6. (a) Proposed multiple feature classification. (b) Multiple feature classification proposed in [18].

different distributions, the information is not equally represented in the stacked vector, and some important features may get lost or mixed if we project stacked features from different sources into a low-dimensional feature space together. We verified experimentally that the first strategy is the more effective one.

C. Cloud-Shadow Detection

Cloud shadow cannot be always avoided during the acquisition of optical remote sensing data. The presence of cloud and shadow complicates the analysis of remote sensing data, leading, e.g., to false detection of land cover change [26], biased estimation of normalized difference vegetation index values, and mistakes in classification tasks. Therefore, the detection of cloud shadow is an initial and important step [35]. Actually, many approaches have been developed to detect cloud shadow, such as geometry-based methods [36] and the Fmask algorithm [37], [38].

Since cloud-shadow detection is not the primary goal of this work, we will apply a simple method based on area attribute filters [31] to detect the big cloud-shadow region, because in our specific case study area, the area of the cloud-shadow is much larger and darker than other ground objects. By increasing the thresholds of the area attribute, more and more bright objects are filtered out, leaving finally the largest dark cloud shadow region [see Fig. 7(b)]. The cloud-shadow mask is then obtained by binarizing the result [see Fig. 7(c)]. In fact, there is a very small cloud-shadow region present at the top center-right of the image. We just choose the large main cloud-shadow as an example here, as all shadowed testing samples are located in this large cloud-shadow region. Denote $\mathbf{G} = \{g_{ij}\}$ as the cloud-shadow mask, with pixel values $g_{ij} = 0$ in the cloud-shadow region and $g_{ij} = 1$ in the shadow-free region.

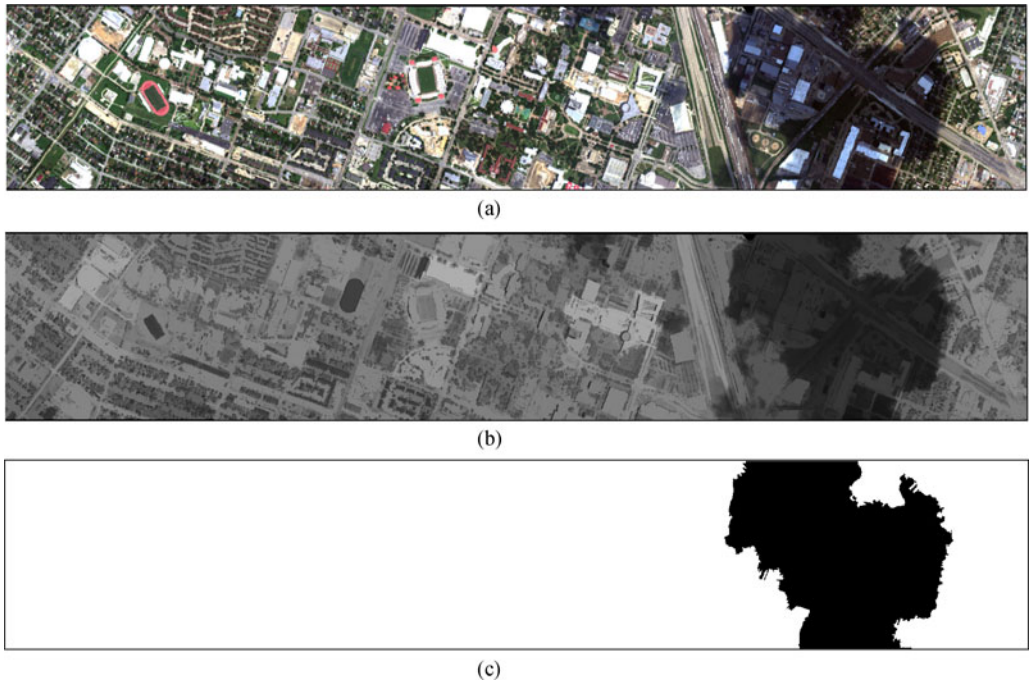


Fig. 7. (a) False RGB image of HS data. (b) Area attribute thinning with area size 3000 of HS data. (c) Extracted cloud-shadow map.

D. Cotraining Samples Selection

In this section, we describe a new method to select and label a separate training set (called cotraining samples) for the cloud-shadow regions. Since LiDAR image are not influenced by clouds, our proposed method uses single elevation information (i.e., Flid, see Section II-B) to obtain an initial classification map (Map_{lid}). However, single elevation information from LiDAR image is not sufficient for a reliable classification, as many man-made objects in urban areas are of similar height. Therefore, we combine the spectral and spatial information from the HS image with Map_{lid} to select new cotraining samples from cloud-shadow regions in the following way [see Fig. 8(a)]:

Suppose $\mathbf{X}^{\text{Sta}} = \{\mathbf{x}_i^{\text{Sta}}\}_{i=1}^{n}$ denotes the set of samples, $\mathbf{x}_i^{\text{Sta}} = \{\mathbf{x}_i^{\text{spe}}; \mathbf{x}_i^{\text{spa}}\}$, $\mathbf{Y}^{'} = \{y_i^{'}\}_{i=1}^{n}$, $\mathbf{x}_i^{\text{spe}}$ and $\mathbf{x}_i^{\text{spa}}$ denote the spectral information in the HS image and spatial information in $\text{EMAPs}_{\text{hsi}}$ of the ith pixel, respectively, $y_i^{'} \in \{1, \ldots, C\}$ denotes the label of pixel i in the classification map [Map_{lid}, see Fig. 8(a)] obtained by the LiDAR feature source.

In fact, multiple feature sources (i.e., original spectrum of HS image, $\text{EMAPs}_{\text{hsi}}$ from HS image, and MAPs_{lid} from LiDAR image) can be seen as information from different aspects for pixels. For two samples, if their information are similar from all aspects, we assume they belong to same class and share same labels. Let $\mathbf{X}_{c(1)}^{'}$ be a set of initial selected cotraining samples which belong to class c. $\mathbf{X}_{c(1)}^{'}$ can be obtained from the following criterion:

$$\mathbf{X}_{c(1)}^{'} = \Big\{\mathbf{x}_i^{Sta} : \mathbf{x}_i^{\text{spe}} \in knn\big(\mathbf{m}_{c(1)}^{\text{spe}}\big) \cdots \wedge \ \mathbf{x}_i^{\text{spa}} \in knn\big(\mathbf{m}_{c(1)}^{\text{spa}}\big)\Big\} \tag{6}$$

where

$$\mathbf{m}_{c(1)}^{\text{spe}} = \frac{1}{n_c^0} \sum_{i=1}^{n_c^0} \mathbf{x}_i^{\text{spe}}, \text{ where } y_i^{'} = c \tag{7}$$

$$\mathbf{m}_{c(1)}^{\text{spa}} = \frac{1}{n_c^0} \sum_{i=1}^{n_c^0} \mathbf{x}_i^{\text{spa}}, \text{ where } y_i^{'} = c \tag{8}$$

"$\wedge$" here means "and," $\mathbf{m}_{c(1)}^{\text{spe}}$ and $\mathbf{m}_{c(1)}^{\text{spa}}$ can be seen as the initial center of class c in spectral feature space and spatial feature space, respectively, n_c^0 is the number of initial selected cotraining samples in class c based on Map_{lid}, $knn(\mathbf{m}_{c(1)})$ denotes the set of k-nearest neighbors of $\mathbf{m}_{c(1)}$. Here k-nearest neighbors are selected based on Euclidean distance, as Euclidean distance is simple and widely used in k-nearest neighbors searching. In this way, the selected cotraining samples for each class have similar spectral, spatial and elevation information. For details see Fig. 8(b).

However, as it is only based on a single elevation feature source, the classification accuracies for some classes are relatively low in Map_{lid}, leading to less accurate class center. As a result, the cotraining samples selected based on (6) are not reliable. In order to solve this problem, we iteratively update the class centers, similar as in a mean shift algorithm. Suppose $\mathbf{X}_{c(k)}^{'}$ is a set of training samples belonging to class c, generated in the kth iteration. This set can be obtained from the training set at the $(k-1)$th iteration through the following criterion:

$$\mathbf{X}_{c(k)}^{'} = \Big\{\mathbf{x}_i^{Sta} : \mathbf{x}_i^{\text{spe}} \in knn\big(\mathbf{m}_{c(k)}^{\text{spe}}\big) \cdots \wedge \ \mathbf{x}_i^{\text{spa}} \in knn\big(\mathbf{m}_{c(k)}^{\text{spa}}\big)\Big\} \tag{9}$$

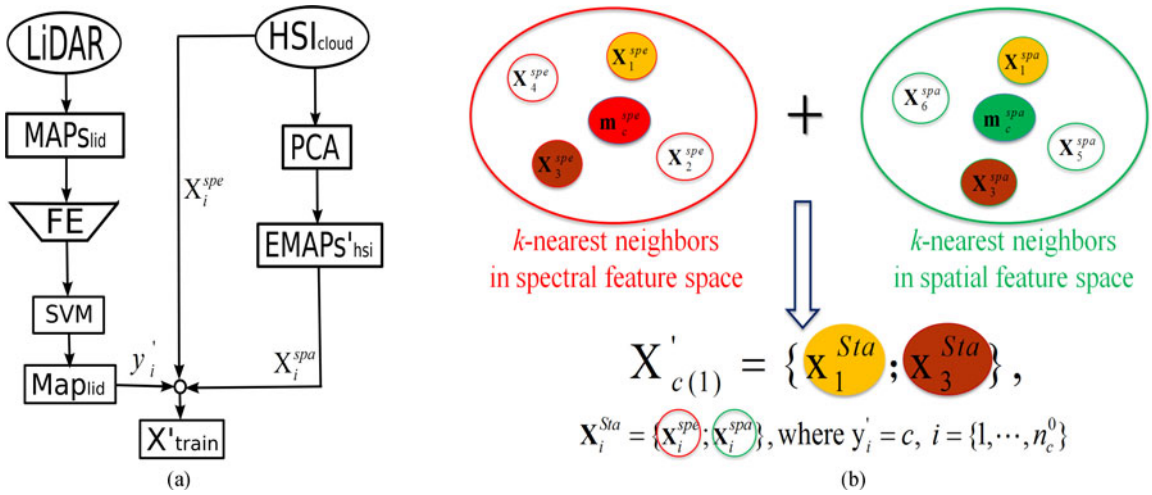


Fig. 8. (a) Cotraining samples selection; $\mathbf{x}_i^{\mathrm{spe}}$ and $\mathbf{x}_i^{\mathrm{spa}}$ represent spectral radiance and $\mathrm{EMAPs}_{\mathrm{hsi}}$ of the ith pixel; (b) $\mathbf{x}_i^{\mathrm{Sta}} = \{\mathbf{x}_i^{\mathrm{spe}}; \mathbf{x}_i^{\mathrm{spa}}\}$; the pixel is labeled c in $\mathrm{Map}_{\mathrm{lid}}$ ($y_i^{'} = c$); $\mathbf{m}_c^{\mathrm{spe}}$ and $\mathbf{m}_c^{\mathrm{spa}}$ are the centers of class c in spectral and spatial feature space, here $\mathbf{x}_1^{\mathrm{Sta}}$ and $\mathbf{x}_3^{\mathrm{Sta}}$ are the nearest neighbors of the center of class c, both in spectral and spatial feature space, and selected as candidate cotraining samples.

where

$$\mathbf{m}_{c(k)}^{\mathrm{spe}} = \frac{1}{n_c^{(k-1)}} \sum_{i=1}^{n_c^{(k-1)}} \mathbf{x}_i^{\mathrm{spe}}, \text{ where } \mathbf{x}_i^{Sta} \in \mathbf{X}_{c(k-1)}^{'} \quad (10)$$

$$\mathbf{m}_{c(k)}^{\mathrm{spa}} = \frac{1}{n_c^{(k-1)}} \sum_{i=1}^{n_c^{(k-1)}} \mathbf{x}_i^{\mathrm{spa}}, \text{ where } \mathbf{x}_i^{Sta} \in \mathbf{X}_{c(k-1)}^{'} \quad (11)$$

$\mathbf{m}_{c(k)}^{\mathrm{spe}}$ and $\mathbf{m}_{c(k)}^{\mathrm{spa}}$ denote the centers of $\mathbf{X}_{c(k-1)}^{'}$ in spectral feature space and spatial feature space, respectively, $n_c^{(k-1)}$ is the number of samples in $\mathbf{X}_{c(k-1)}^{'}$. The iteration procedure can be stopped by introducing two thresholds $\varepsilon^{\mathrm{spe}}$ and $\varepsilon^{\mathrm{spa}}$. When

$$\left|\mathbf{m}_{c(k)}^{\mathrm{spe}} - \mathbf{m}_{c(k-1)}^{\mathrm{spe}}\right| < \varepsilon^{\mathrm{spe}} \wedge \left|\mathbf{m}_{c(k)}^{\mathrm{spa}} - \mathbf{m}_{c(k-1)}^{\mathrm{spa}}\right| < \varepsilon^{\mathrm{spa}} \quad (12)$$

the centers of cotraining samples in each class are stable. We define the final cotraining samples set as: $\mathbf{X}_{\mathrm{train}}^{'} = \{\mathbf{X}_{1(k)}^{'}, \mathbf{X}_{2(k)}^{'}, \ldots, \mathbf{X}_{c(k)}^{'}\}$. The algorithmic procedure of the proposed cotraining samples selection method is formally stated in Algorithm 1.

Algorithm 1: Co-training samples selection algorithm.

1: **Input:** Samples under cloud shadow $\mathbf{X}^{Sta} = \{\mathbf{x}_i^{Sta}\}_{i=1}^{n}$, $\mathbf{x}_i^{Sta} = \{\mathbf{x}_i^{\mathrm{spe}}; \mathbf{x}_i^{\mathrm{spa}}\}$, and their labels $\mathbf{Y}^{'} = \{y_i^{'}\}_{i=1}^{n}$ in $\mathrm{Map}_{\mathrm{lid}}$.
2: Calculate the initial spectral center of every class $\mathbf{m}_{c(1)}^{\mathrm{spe}} (c \in \{1, \cdots, C\})$ via (7).
3: Calculate the initial spatial center of every class $\mathbf{m}_{c(1)}^{\mathrm{spa}} (c \in \{1, \cdots, C\})$ via (8).
4: Find the common nearest neighbors $\mathbf{X}_{c(1)}^{'}$ via (6).
5: $k = 2$
6: **Loop**
7: Update $\mathbf{m}_{c(k)}^{\mathrm{spe}}$ via (10).
8: Update $\mathbf{m}_{c(k)}^{\mathrm{spa}}$ via (11).
9: **if** ($|\mathbf{m}_{c(k)}^{\mathrm{spe}} - \mathbf{m}_{c(k-1)}^{\mathrm{spe}}| < \varepsilon^{\mathrm{spe}}$ and $|\mathbf{m}_{c(k)}^{\mathrm{spa}} - \mathbf{m}_{c(k-1)}^{\mathrm{spa}}| < \varepsilon^{\mathrm{spa}}$) **then**
10: **break Loop**
11: **end if**
12: Update $\mathbf{X}_{c(k)}^{'}$ via (9).
13: $k \leftarrow k + 1$
14: **End Loop**
15: **Output:** Final selected co-training samples $\mathbf{X}_{\mathrm{train}}^{'} = \{\mathbf{X}_{1(k)}^{'}, \mathbf{X}_{2(k)}^{'}, \cdots, \mathbf{X}_{C(k)}^{'}\}$.

E. Classification Map Fusion

After obtaining the new cotraining samples under cloud-shadow regions, multiple features classification is applied in the same way as for cloud-free regions, the only difference being the way the training samples were obtained, where in the cloud-free regions, we use available training samples outside of the cloud mask, while in the cloud-shadow regions, we apply our proposed cotraining samples selection procedure. The final classification map is obtained by the fusion of the two maps: $\mathrm{Map}_{\mathrm{cloud}}$ and $\mathrm{Map}_{\mathrm{nocloud}}$

$$\mathrm{Map} = g_{i,j}\mathrm{Map}_{\mathrm{cloud}} + \overline{g}_{i,j}\mathrm{Map}_{\mathrm{nocloud}} \quad (13)$$

where $\overline{g}_{i,j}$ is the logical inverse of $g_{i,j}$.

III. Experiments

A. Data Description

In 2013, the Data Fusion Technical Committee of the IEEE GRSS organized a contest involving two types data sources: a cloud-shadow HS image and a LiDAR-derived DSM, both at the same spatial resolution (2.5 m) [20]. The competition was established to devise advanced methods for the fusion and classification of HS and LiDAR data [21]. This dataset was captured by the NSF-funded Center for Airborne Laser Mapping (NCALM) using the compact airborne spectrographic imager (CASI-1500)

TABLE II
CLASSIFICATION ACCURACIES FOR THE SHADOW-FREE REGIONS OBTAINED BY THE DIFFERENT METHODS

Features	Raw_{hsi}	$Fspe$	$Fspa$	$Flid$	$FEstacked$ [18]	$Ffusion$	$Fggf$ [22]	Proposed
Number of Features	144	15	15	15	45	45	22	45
Grass Healthy (198/875)	98.86	98.40	99.54	69.71	99.66	**100.00**	99.77	**100.00**
Grass Stressed (190/906)	96.36	96.58	89.51	71.52	97.90	98.34	**99.34**	98.34
Grass Synthetis (192/505)	99.80	**100**	**100**	94.85	**100**	**100**	**100**	**100**
Tree (188/986)	98.49	95.54	89.55	74.54	98.99	97.26	**99.19**	97.26
Soil (186/1056)	97.92	98.58	98.30	83.14	99.43	**99.83**	99.81	**99.83**
Water (182/143)	95.10	95.10	**99.30**	84.62	95.80	95.80	95.80	95.80
Residential (196/992)	82.36	79.44	83.17	84.58	92.14	**98.99**	94.46	**98.99**
Commercial (191/622)	72.67	82.80	63.50	92.44	93.73	**93.89**	91.64	**93.89**
Road (193/1040)	79.35	76.66	80.50	66.28	86.74	**94.43**	89.34	**94.43**
Highway (191/710)	86.62	93.66	96.06	93.52	96.34	**99.72**	91.13	**99.72**
Railway (181/902)	91.57	90.47	87.58	96.78	96.78	**99.22**	95.01	**99.22**
Parking Lot 1 (192/1041)	84.34	78.19	88.18	66.38	89.15	**97.98**	81.08	**97.98**
Parking Lot 2 (184/265)	77.36	76.23	78.11	68.68	86.04	**87.92**	80.00	**87.92**
Tennis Court (181/247)	99.60	99.60	**100**	99.60	**100**	**100**	**100**	**100**
Running Track (187/473)	97.25	97.04	**100**	58.35	98.52	98.73	98.73	98.73
OA (%)	90.28	89.77	89.42	79.03	95.24	**97.91**	94.37	**97.91**
AA (%)	90.58	90.55	90.22	80.33	95.37	**97.47**	94.35	**97.47**
κ	0.894	0.889	0.885	0.773	0.948	**0.977**	0.939	**0.977**

on June 2012 over the University of Houston campus and its neighboring urban area. The HS image has 144 spectral bands ($D = 144$) with a wavelength range from 380 to 1050 nm. The whole scene of the data contains 349 $\times$ 1905 pixels. The ground truth provided for this dataset contains 15 classes, summed up in Table II, also mentioning between brackets the available numbers of training/test samples. The false color image and LiDAR image are shown in Fig. 1(a) and (b), the distribution of training and test samples are shown in Fig. 1(c) and (d). The given scene contains a large cloud-shadow region [see Fig. 1(a)], which distorts the spectral radiance of objects in the HS image (darkening effect). More information can be found in [20].

B. Experimental Setup

The input HS image is transformed by PCA, and the first two PCs are kept since they contain almost all of the variance in the dataset (cumulative variation of more than 99%). For the FE, we use a supervised method: NWFE [34], as it has been shown to be efficient in many classification applications [32]. To generate the EMAPs, four attributes are considered:

1) "*a*" area λ_a (related the size of the objects);
2) "*s*" standard deviation λ_s (as a measure of homogeneity of the objects);
3) "*d*" diagonal of the box bounding the objects λ_d;
4) "*i*" moment of inertia λ_i (as a measure of the elongation of the objects).

The values of each attribute are given as follows:

λ_a = [50 100 200 300 500 700 1000 1500 2000 2500 3000 4000];

λ_s = [5 10 15 20 25 30 35 40 50 60];

λ_d = [5 10 25 50 75 100 150 200 300 400 500];

λ_i = [0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1].

According to the number of selected values of each attribute, the dimension (denoted by D_1) of $\text{EMAPs}_{\text{hsi}}$ is $D_1 = 134$, the dimension (denoted by D_2) of MAPs_{lid} is $D_2 = 67$.

The SVM classifier [39] with radial basis function (RBF) kernels is applied, containing two parameters: the penalty factor C and the RBF kernel widths γ. C is optimized within the given set $\{10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ and γ is optimized within the given set $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1\}$ by fivefold cross validation.

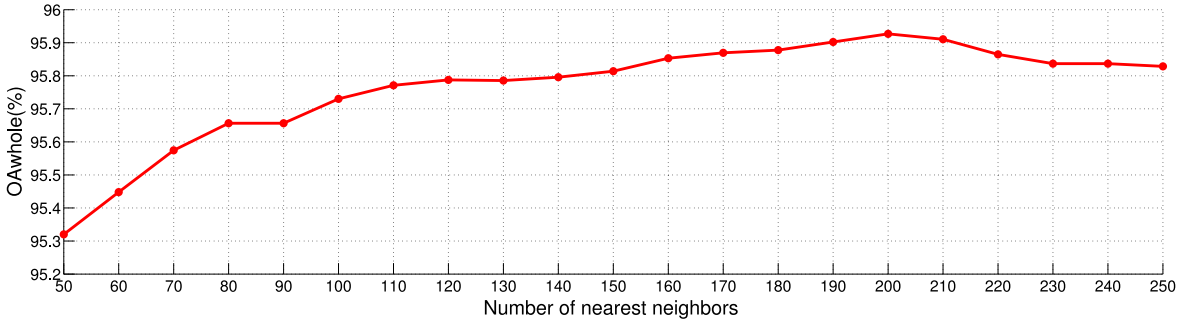
We compare our proposed framework with the following schemes:

1) using the original HS image (Raw_{hsi});
2) using spectral features Fspe extracted from the HS image by NWFE;
3) using spatial features Fspa extracted from $\text{EMAPs}_{\text{hsi}}$ by NWFE;
4) using elevation features Flid extracted from MAPs_{lid} by NWFE;
5) using stacked features (FEstacked), stacking all spectral features, $\text{EMAPs}_{\text{hsi}}$ and MAPs_{lid} first, similar to the approach of [18], and then extracting low-dimensional features by NWFE, as shown in Fig. 6(b);
6) using fusion of stacked features Fspe, Fspa, and Flid but only using the original training samples (Ffusion); this is the proposed approach without the cotraining samples selection procedure.
7) using fused features from the generalized graph-based fusion method (Fggf), the same as in the approach of [22].

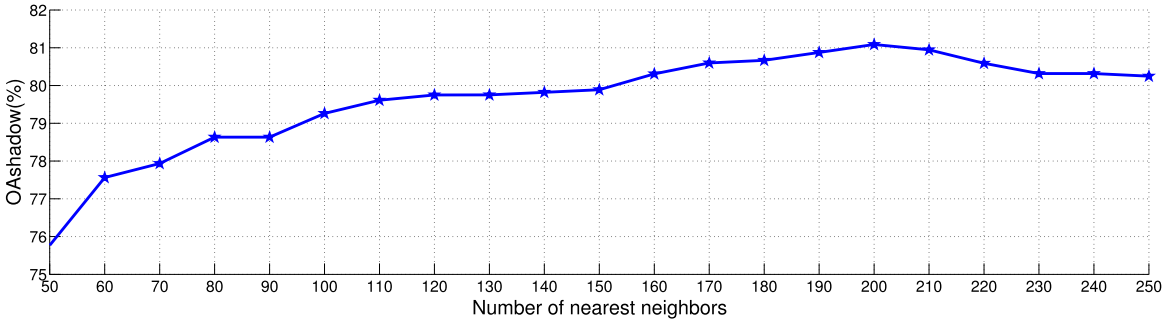
The classification results are quantitatively evaluated by measuring the overall classification accuracy (OA), the average accuracy (AA), Kappa coefficient (κ) on the test samples, as shown in Fig. 1(d).

C. Effect of Number of Nearest Neighbors for Cotraining Selection

The number of nearest neighbors of each class center (e) is an important parameter in the cotraining samples selection procedure. On the one hand, when e is too small, there will be insufficient cotraining samples. On the other hand, large e will lead to mislabeling of cotraining samples, as some samples with different labels will be included in the nearest neighbors. To investigate the effect of the number of nearest neighbors on the classification accuracy, we performed classification experiments



(a)



(b)

Fig. 9. (a) OA on the whole scene with increasing number of nearest neighbors. (b) OA on shadow area with increasing number of nearest neighbors.

TABLE III
CLASSIFICATION ACCURACIES FOR THE CLOUD-SHADOW REGIONS OBTAINED BY THE DIFFERENT METHODS

Features	Raw$_{hsi}$	$Fspe$	$Fspa$	$Flid$	$FEstacked$ [18]	$Ffusion$	$Fggf$ [22]	Proposed
Number of Features	144	15	15	15	45	45	22	45
Grass Healthy (85/178)	0.00	0.00	0.00	55.06	0.00	0.00	0.00	**84.26**
Grass Stressed (102/158)	0.00	0.00	0.00	51.89	0.00	0.00	**91.78**	89.24
Tree (89/70)	0.00	0.00	0.00	**100.00**	0.00	0.00	**100.00**	**100.00**
Residential (75/80)	1.25	71.25	**90.00**	58.75	0.00	66.25	63.74	77.50
Commercial (121/431)	29.23	0.00	0.00	82.60	69.37	49.19	84.74	**88.17**
Road (71/19)	0.00	0.00	0.00	**63.16**	0.00	0.00	52.63	57.89
Highway (92/326)	0.00	51.23	0.00	46.63	15.34	23.93	**76.07**	69.49
Railway (111/152)	4.61	71.05	15.79	92.11	91.45	92.11	**93.42**	92.76
Parking Lot 2 (27/20)	5.00	**10.00**	0.00	5.00	0.00	0.00	0.00	**10.00**
OA (%)	9.42	22.31	6.70	66.71	34.96	34.54	71.97	**81.15**
AA (%)	4.45	22.61	11.75	61.56	20.52	36.60	62.49	**74.37**
κ	0.160	0.134	0.111	0.679	0.206	0.230	0.717	**0.796**

with different numbers of e. The number of nearest neighbors was changed from 50 to 250 with a stepsize of 10. Fig. 9 shows the OA for the whole scene and for the shadow area in the function of an increasing number of nearest neighbors. As can be seen, the average OA increases as the number of nearest neighbors grows from 50 to 200, and then decreases with more nearest neighbors. This indicates that if e is set to a small value, the selected number of cotraining samples will be too small to allow us to train the classifier; if e is set to a large value, the possibility of mislabeling cotraining samples increases, leading to poor classification performances. For this dataset, we have set the number of nearest neighbors to 200 in our experiments.

D. Classification Results on the Dataset

This section mainly explores the efficiency of the proposed method, compared with the other methods. The resulting accuracies are reported in Tables II–IV, and the classification maps are shown in Fig. 10 for visual comparison. The best accuracy among different methods (in row) is highlighted in bold. From the tables and figures, we conclude the following.

1) The proposed framework improves all results in terms of the OA, the AA, the Kappa coefficient (κ), and the quality of the classification map. On the shadow-free region, it outperforms the state of the art by 2%, in the cloud-shadow region, the improvements are dramatic. On the whole scene, the proposed framework improves the OA with 3.87–20.10% over the other schemes.
2) In general, it can be observed that fusion of multiple features (spectral, spatial, and elevation features) leads to better classification performances in comparison with using one single type of features. This shows that the chosen sets of features are efficient and fusing them exploits the information contained in both data sources.

TABLE IV
CLASSIFICATION ACCURACIES FOR THE WHOLE IMAGE OBTAINED BY THE DIFFERENT METHODS

Features	Raw$_{hsi}$	$Fspe$	$Fspa$	$Flid$	$FEstacked$ [18]	$Ffusion$	$Fggf$ [22]	Proposed
Number of Features	144	15	15	15	45	45	22	45
Grass Healthy (1053)	82.15	81.77	82.72	57.93	82.81	83.00	82.91	**97.34**
Grass Stressed (1064)	82.05	82.24	76.22	60.90	83.36	83.74	**99.44**	96.99
Grass Synthetis (505)	99.80	**99.80**	**100**	94.85	**100**	**100**	**100**	**100**
Tree (1056)	92.90	89.20	83.62	76.23	92.42	91.48	**99.24**	97.44
Soil (1056)	97.92	98.58	98.30	83.14	99.43	99.81	99.81	**99.83**
Water (143)	95.10	95.10	**99.30**	84.62	95.80	95.80	95.80	95.80
Residential (1072)	76.31	78.82	83.68	82.65	85.26	97.11	91.42	**97.38**
Commercial (1053)	54.89	48.91	37.51	**92.97**	83.76	75.59	92.50	91.55
Road (1059)	78.00	75.35	79.13	66.48	85.27	92.82	88.76	**93.86**
Highway (1036)	59.36	80.31	65.83	70.17	70.85	75.87	84.85	**90.15**
Railway (1054)	79.03	87.67	77.23	96.39	97.25	**99.15**	95.73	98.58
Parking Lot 1 (1041)	84.34	78.19	88.18	66.38	89.15	97.69	81.08	**97.69**
Parking Lot 2 (285)	72.28	71.58	72.63	63.86	80.00	77.89	74.39	**82.46**
Tennis Court (247)	99.60	99.60	**100**	99.60	**100**	**100**	**100**	**100**
Running Track (473)	97.25	97.04	**100**	58.35	98.52	98.73	98.73	98.73
OA (%)	80.78	81.96	79.70	75.82	88.16	90.44	92.05	**95.92**
AA (%)	83.40	84.29	82.96	76.97	89.55	91.25	92.31	**95.65**
κ	0.793	0.805	0.779	0.738	0.871	0.896	0.914	**0.958**

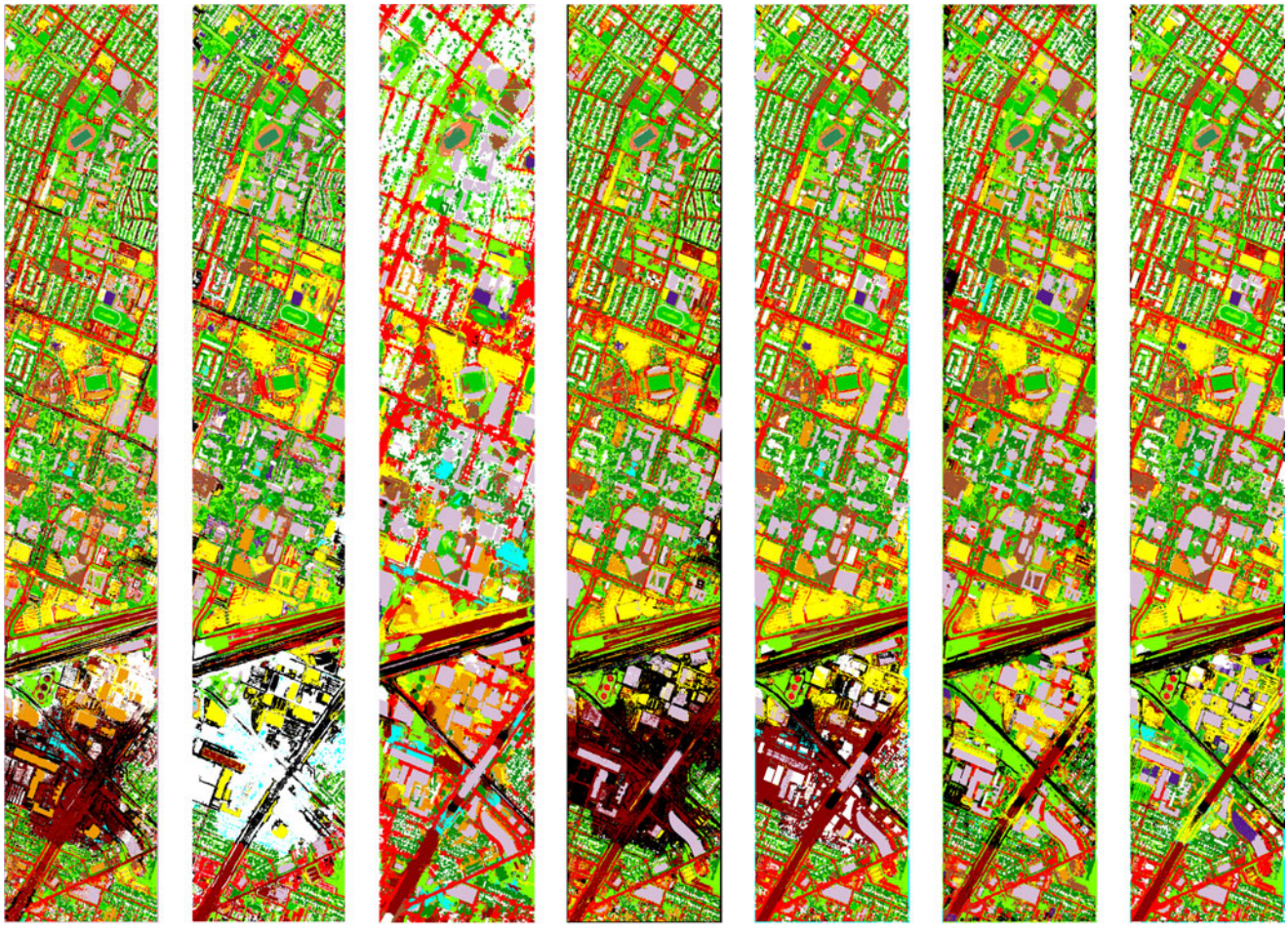


Fig. 10. Classification maps produced by the different methods. (a) Fspe; (b) Fspa; (c) Flid; (d) FEstacked; (e) Ffusion; (f) Fggf; and (g) Proposed.

3) When investigating the classification accuracies for each class separately in Table IV, it can be clearly noticed that, when single features are used, the Raw_{hsi} approach produces better results on class "Tree," whereas the Fspa scheme performs better on classes "water," "Residential" and "Road." However, spectral or spatial features from the HS image perform poor on classes "Commercial" and "Railway." On the contrary, Flid, extracted from the LiDAR image performs much better on these two classes. Classification accuracies for most classes improve by fusing those three features, especially for classes "Residential," "Road," and "Parking Lot 2." The generalized graph-based fusion method (Fggf) [22] improves the classification accuracy on classes "Grass Stressed," "Tree," and "Highway." The proposed framework obtains the best classification accuracies on 9 of the 15 classes.
4) From the results reported on the shadow-free region (see Table II) and the whole image (see Table IV), one can infer that fusing the features extracted from each source (HS image, $\text{EMAPs}_{\text{hsi}}$, and MAPs_{lid}) works better than using the features extracted from the stacked vector of the original HS image, $\text{EMAPs}_{\text{hsi}}$, and MAPs_{lid}, with an improvement of almost 3%. The main reason behind this is that, because of their different nature, when fusing features from different sources and then projecting them on a lower dimensional space, information gets mixed up and lost.
5) By comparing the classification maps in Fig. 10 and classification accuracies on the cloud-shadow region (see Table III), we can see that most of the objects under the cloud-shadow region are not well classified when only using the training samples located in the shadow-free region. Some objects in the cloud-shadow region are classified better by using features extracted from LiDAR image, because the elevation information contained in the morphological features of LiDAR image is not influenced by the cloud. But for many other objects, the results are not good as the elevation information is not sufficiently discriminative. Taking all feature sources into consideration does not much improve the classification accuracy for most of the classes. The proposed framework leads to an improved classification of most classes due to the selection and use of specific training samples in the cloud-shadow region.

Moreover, as the described dataset [20] is very popular and open access, it has been used in many recently state of the arts, such as in [18] and [22]–[24]. Compared with the experimental results from these references, the proposed scheme performs better on either cloud-shadow or shadow-free regions, with overall classification accuracy 97.91% and 81.15%, respectively. Thus proves the proposed fused features are effective and distinguishable, and selecting new training samples from the cloud-shadow regions is an efficient solution.

IV. Conclusion

In this paper, we developed a new method for the classification of cloud mixed remote sensing scenes by the fusion of HS and LiDAR data. The proposed method selects new samples as cotraining samples in the cloud-shadow regions and classifies shadow-free and cloud-shadow regions separately using their own sets of training samples. In order to better combine HS and LiDAR data, additional spatial and elevation features are extracted from HS and LiDAR data, and effectively integrated without any regularization or weight parameters. Experimental results on the classification of the real cloud-shadow HS and LiDAR data show the efficiency of the proposed framework. In addition, the proposed approach can be thought of as a general framework, in which the FE step can be replaced by any other technique (kernel PCA, supervised or semisupervised FE, ...), possibly to improve classification accuracies. Moreover, the proposed framework is completely open and flexible in its capacity to integrate more features.

Recent Earth observation missions (Landsat series from NASA, Sentinel series from ESA) boost the use of the multisensor remote sensing imagery. However, cloud/shadow effects cannot be avoided in the optical sensors. Other sensors (e.g., synthetic aperture radar, thermal infrared, LiDAR, etc.) can provide complementary information for these cloud/shadow regions. The proposed framework is applicable for the fusion of optical HS images and other images (e.g., SAR, thermal infrared), where multisensor images are available. On the other hand, deshadowing is typically used as a preprocessing step before applications. Fusion of multisensor data for deshadowing (e.g., HS image restoration) will be our future work.

Acknowledgements

The authors would like to thank the Hyperspectral Image Analysis group and the NCALM at the University of Houston for providing the data sets used in this study.

References

[1] D. Bruce *et al.*, "NASA Goddard's LiDAR, hyperspectral and thermal (G-LiHT) airborne imager," *Remote Sens.*, vol. 5, no. 8, pp. 4045–4066, Aug. 2013.
[2] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.
[3] R. Luo, W. Liao, X. Huang, Y. Pi, and W. Philips., "Feature extraction of hyperspectral images with semi-supervised graph learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4389–4399, Sep. 2016.
[4] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.
[5] J. Jung, E. Pasolli, S. Prasad, J. Tilton, and M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 491–502, Feb. 2014.
[6] 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. 4, pp. 5312–5326, Oct. 2015.
[7] A. F. Elakshe, "Fusion of hyperspectral images and LiDAR-based DEMs for coastal mapping," *Opt. Lasers Eng.*, vol. 46, no. 7, pp. 493–498, Jul. 2008.
[8] 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.

[9] 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.

[10] L. Naidooa, M. Choa, R. Mathieua, 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.

[11] M. Shimoni, G. Tolt, C. Perneel, and J. Ahlberg, "Detection of vehicles in shadow areas," in *Proc. 3rd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2011, pp. 1–4.

[12] M. Pedergnana, P. Reddy Marpu, M. Dalla 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.

[13] W. Liao, R. Bellens, A. Pižurica, S. Gautama, and W. Philips, "Combining feature fusion and decision fusion for classification of hyperspectral and LiDAR Data," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Quebec City, QC, Canada, Jul. 2014, vol. 1, pp. 1241–1244.

[14] C. Chen, W. Li, H. Su, and K. Liu, "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sens.*, vol. 6, no. 6, pp. 5795–5814, Jun. 2014.

[15] P. Ghamisi, J. Benediktsson, and M. Ulfarsson, "Spectral-spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, May 2014.

[16] W. Liao, R. Bellens, A. Pižurica, W. Philips, and Y. Pi, "Classification of hyperspectral data over urban areas using directional morphological profiles and semi-supervised feature extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1177–1190, Aug. 2012.

[17] M. Dalla 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, Nov. 2010.

[18] 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.

[19] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55–63, Jan. 1968.

[20] 2013 IEEE GRSS Data Fusion Contest. [Online]. Available: http://www.grssieee.org/community/technical-committees/data-fusion/

[21] 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.

[22] W. Liao, A. Pižurica, 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.

[23] R. Bao, J. Xia, M. Dalla Mura, P. Du, J. Chanussot, and J. Ren, "Combining morphological attribute profiles via an ensemble method for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 359–363, Mar. 2016.

[24] Z. Zhong, B. Fan, K. Ding, H. Li, S. Xiang, and C. Pan, "Efficient multiple feature fusion with hashing for hyperspectral imagery classification: A comparative study," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4461–4478, Aug. 2016.

[25] P. Ghamisi, B. Hofle, and X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published. doi: 10.1109/JSTARS.2016.2634863.

[26] Z. Zhu and C. E. Woodcock, "Cloud, cloud shadow, and snow detection in multitemporal landsat data: An algorithm designed specifically for monitoring land cover change," *Remote Sens. Environ.*, vol. 152, pp. 217–234, Sep. 2014.

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

[28] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.

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

[30] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[31] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 541–545, May 2011.

[32] P. Ghamisi, J. Benediktsson, and J. Sveinsson, "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.

[33] 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. 40, no. 3, pp. 480–491, Mar. 2005.

[34] B. Kuo and D. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.

[35] T. Arvidson, J. Gasch, and S. N. Goward, "Landsat-7's long-term acquisition plan-An innovative approach to building a global imagery archive," *Remote Sens. Environ.*, vol. 78. nos. 1/2, pp. 13–26, Oct. 2001.

[36] Y. Luo, A. P. Trishchenko, and K. V. Khlopenkov, "Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America," *Remote Sens. Environ.*, vol. 112, no. 12, pp. 4167–4185, Dec. 2008.

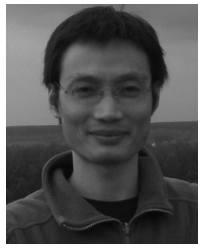
[37] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, no. 15, pp. 83–94, Mar. 2012.

[38] D. Frantz, A. Roder, T. Udelhoven, and M. Schmidt, "Enhancing the detectability of clouds and their shadows in multitemporal dryland Landsat imagery: Extending fmask," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1242–1246, Jun. 2015.

[39] C. Chang and C. Lin, "A library for support vector machines." Dec. 2016. [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm

Renbo Luo (S'16) received the B.S. degree in Automation Control Engineering from South China University of Technology, Guangzhou, China, in 2010. Currently, he is working toward the Ph.D. degree in Control Theory and Control Engineering from South China University of Technology, China, and the Ph.D. Degree in Computer Science Engineering from Ghent University, Belgium.

His research interests include image processing, object recognition, data fusion, domain adaptation, and remote sensing applications



Wenzhi Liao (S'10–M'14–SM'16) received the B.S. degree in mathematics from Hainan Normal University, Haikou, China, in 2006, the Ph.D. degree in engineering from 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 working as a Postdoc at Ghent University and as a Postdoctoral Fellow with the Fund for Scientific Research, Flanders-FWO. His current research interests include pattern recognition, remote sensing, and image processing. In particular, his interests include mathematical morphology, multitask feature learning, multisensor data fusion, and hyperspectral image restoration.

Dr. Liao is a member of the Geoscience and Remote Sensing Society (GRSS) and IEEE GRSS Data Fusion Technical Committee (DFTC). He received the "Best Paper Challenge" Awards on both 2013 IEEE GRSS Data Fusion Contest and 2014 IEEE GRSS Data Fusion Contest.

Hongyan Zhang (M'13–SM'16) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

Since 2013, he has been an Associate Professor in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He has authored/coauthored more than 60 research papers. His research interests include image reconstruction for quality improvement, hyperspectral image processing, sparse representation. and low rank methods for sensing image imagery.

Dr. Zhang is a reviewer of more than 20 international academic journals, including IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTION ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, etc.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 500 research papers and five books. He is the holder of 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He is a Fellow of the Institution of Engineering and Technology, an Executive Member (Board of Governors) of the China National Committee of the International GeosphereBiosphere Programme, an Executive Member of the China Society of Image and Graphics, etc. He regularly serves as a Co-Chair of the series International Society for Optics and Photonics (SPIE) Conferences on Multispectral Image Processing and Pattern Recognition, the Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, the *Journal of Remote Sensing*, and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING; he is a Guest Editor of the *Journal of Applied Remote Sensing* and the *Journal of Sensors*. He was the General Chair for the Fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He received the Best Reviewer Awards from the IEEE GRSS for his service to the IEEE JSTARS in 2012 and the IEEE Geoscience and Remote Sensing Letters in 2014. He also received the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing. His research teams received the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium Student Paper Contest in recent years.

Paul Scheunders (M'98–SM'14) received the B.S. and Ph.D. degrees in physics, with work in the field of statistical mechanics, from the University of Antwerp, Antwerp, Belgium, in 1983 and 1990, respectively.

In 1991, he became a Research Associate with the Vision Lab, Department of Physics, University of Antwerp, where he is currently a Professor. He has published more than 200 papers in international journals and proceedings in the field of image processing, pattern recognition, and remote sensing. His current research interest includes remote sensing and hyperspectral image processing.

Dr. Scheunders is an Associate Editor of the IEEE TRANSACTIONS IN GEOSCIENCE AND REMOTE SENSING, and has served as a Program Committee Member in numerous international conferences on remote sensing. He is Senior Member of the IEEE Geoscience and Remote Sensing Society.

Youguo Pi received the Diploma degree in automation engineering from Chongqing University China and the Ph.D. degree in mechanical engineering from South China University of Technology, Guangzhou, China, in 1982 and 1998, respectively.

From July 1998 to June 2002, he was with the Department of Information Technology, Automation Engineering Center, Academy Guangdong Province. Since July 2002, he has been with the College of Automation Science and Engineering, South China University of Technology, where he is currently a Full-Time Professor. Some of the recent research activities in the group, including image processing and patter recognition and motion control. Important application areas targeted by the group include Intelligent Chinese character formation and servo system.

Wilfried Philips (S'90–M'93–SM'10) was born in Aalst, Belgium, on October 19, 1966. He received the Diploma degree in electrical engineering and the Ph.D. degree in applied sciences, both from Ghent University, Ghent, Belgium, in 1989 and 1993, respectively.

Since November 1997, he has been a Lecturer in the Department of Telecommunications and Information Processing, Ghent University. He is currently a Senior Full Professor in the Department of Telecommunications and Information Processing, Ghent University, where he heads the research group "Image Processing and Interpretation." His main research interests include the domains of image and video restoration and multicamera computer vision.