CoSpace: Common Subspace Learning From Hyperspectral-Multispectral Correspondences

Danfeng Hong, *Student Member, IEEE*, Naoto Yokoya, *Member, IEEE*, Jocelyn Chanussot, *Fellow, IEEE*, and Xiao Xiang Zhu, *Senior Member, IEEE*

***Abstract*— With a large amount of open satellite multispectral (MS) imagery (e.g., Sentinel-2 and Landsat-8), considerable attention has been paid to global MS land cover classification. However, its limited spectral information hinders further improving the classification performance. Hyperspectral imaging enables discrimination between spectrally similar classes but its swath width from space is narrow compared to MS ones. To achieve accurate land cover classification over a large coverage, we propose a cross-modality feature learning framework, called common subspace learning (CoSpace), by jointly considering subspace learning and supervised classification. By locally aligning the manifold structure of the two modalities, CoSpace linearly learns a shared latent subspace from hyperspectral-MS (HS-MS) correspondences. The MS out-of-samples can be then projected into the subspace, which are expected to take advantages of rich spectral information of the corresponding hyperspectral data used for learning, and thus leads to a better classification. Extensive experiments on two simulated HS-MS data sets (University of Houston and Chikusei), where HS-MS data sets have tradeoffs between coverage and spectral resolution, are performed to demonstrate the superiority and effectiveness of the proposed method in comparison with previous state-of-the-art methods.**

***Index Terms*— Common subspace learning (CoSpace), cross-modality learning, hyperspectral, landcover classification, multispectral (MS), remote sensing.**

I. Introduction

RECENTLY, the launch of operational optical broadband [multispectral (MS)] satellites has successfully boosted the usage of MS data for various tasks such as urban monitoring, management of natural resources, ecosystem, and disasters prediction. There has been a growing interest in large-scale land cover mapping of urban [1], agriculture monitoring [2], [3], and mineral exploration [4], since high-quality MS satellite imagery is openly available on a global scale (e.g., Sentinel-2 and Landsat-8). However, MS data fail to discriminate spectrally similar classes due to its broad spectral bandwidth. Hyperspectral imaging can acquire richer spectral information that enables high discrimination ability but its coverage from space is much narrower than the one of MS imaging due to the limitations of imaging devices and satellite techniques. This tradeoff naturally motivates us to ponder a question: *can HS imagery covering only a limited part of the MS imagery be explored to improve the classification of the entire area covered by the MS imagery?* This is as a typical cross-modal feature learning problem.

Researchers have proposed a variety of multimodal feature learning algorithms by introducing additional information, which can be roughly categorized into two parts: fusion-based joint feature learning (FJFL) [5], [6] and alignment-based shared feature learning (ASFL) [7]. The main difference between FJFL and ASFL is illustrated in Fig. 1.

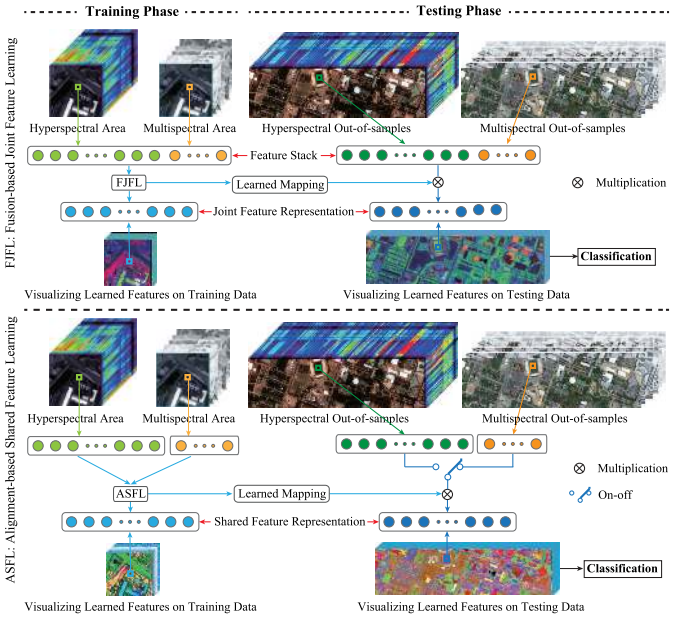


Fig. 1. Illustration of the two kinds of multimodal feature learning frameworks, where the switch (on–off) means that only one modality is involved as the testing samples to meet the hypothesis of cross-modal learning.

Manuscript received April 18, 2018; revised October 11, 2018; accepted December 29, 2018. Date of publication March 20, 2019; date of current version June 24, 2019. This work was supported in part by the European Research Council through the European Union's Horizon 2020 Research and Innovation Programme under Grant ERC-2016-StG-714087, in part by Helmholtz Association through the Framework of the Young Investigators Group "Signal Processing in Earth Observation (SiPEO)" under Grant VH-NG-1018, and in part by the German Research Foundation (DFG) under Grant ZH 498/7-2. The work of N. Yokoya was supported in part by the Japan Society for the Promotion of Science under Grant Grants-in-Aid for Scientific Research (KAKENHI) 15K20955 and in part by Alexander von Humboldt Fellowship for Post-Doctoral Researchers. *(Corresponding author: Xiao Xiang Zhu.)*

D. Hong and X. X. Zhu are with the Remote Sensing Technology Institute (IMF), German Aerospace Center, 82234 Weßling, Germany, and also with Signal Processing in Earth Observation, Technical University of Munich, 80333 Munich, Germany (e-mail: danfeng.hong@dlr.de; xiaoxiang.zhu@dlr.de).

N. Yokoya is with Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project, RIKEN, Tokyo 103-0027, Japan (e-mail: naoto.yokoya@riken.jp).

J. Chanussot is with the GIPSA-Lab, CNRS, Grenoble INP, Université Grenoble Alpes, F-38000 Grenoble, France, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: jocelyn@hi.is).

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/TGRS.2018.2890705

FJFL aims to learn discriminative features by absorbing the different properties from multimodal data. FJFL fuses the different sources at the data level to diversify the information and then to further learn the higher level feature representation. One intuitive way for FJFL is to directly learn a joint data representation at the feature level. At present, this is the mainstream approach for multimodal data analysis [8]. For example, by embedding the height information from light detection and ranging (LiDAR) into MS (HS) data, Ghamisi *et al.* [9] learned multifold features from HS and LiDAR correspondences for a multimodal classification task. Iyer *et al.* [10] provided a graph-based new perspective for feature extraction and segmentation of multimodal images and achieved a desirable result. The resulting discriminative features are beneficial for improving the performance of some high-level applications, especially classification [11], [12], object detection [13], image/video analysis [14], and spectral unmixing [15]. Image fusion can also be regarded as a part of FJFL when feature learning is applied subsequently. For instance, hyperspectral and MS (HS-MS) data fusion enhances the spectral resolution of MS data by fusing it with the low-spatial-resolution HS data [5]. The fused HS-MS product can be then seen as a new input for further discriminative feature learning.

Behind the advancement of FJFL, the *complete data correspondence* is the prerequisite. This limitation undoubtedly results in a poor fit for cross-modal data analysis, in particular, for cross-modal feature learning [16].[1] In our MS-HS case, the cross-modal learning refers to a problem that given a large-scale MS image and a limited HS area partially overlapping with the MS data (see Fig. 2, for example), we learn the low-dimensional embedding representation from the limited amount of MS-HS correspondences and transfer the learned features to the rest of MS data for improving the performance of large-scale land-cover and land-use mapping. During the process, we expect to transfer the discrimination capability learned from the rich spectral information into MS data through the learned common subspace in order to more effectively identify some challenging classes that are hardly recognized by MS data due to its poor spectral information. Please note that we just start a preliminary investigation of cross-modal learning (MS-HS) in this paper, that is, the MS and HS images share the same land-cover classes.

Unlike FJFL, ASFL is more apt for cross-modal feature learning, since ASFL can adaptively shuttle back and forth between the different modalities or domains by means of the learned common subspace. Matasci *et al.* [17] linearly projected the hyperspectral data of the source and target domains into a common feature space where the gap between domains in hyperspectral image classification is expected to be reduced. Kulis *et al.* [18] addressed the issue of visual domain adaption by learning a nonlinear transformation in kernel space, with the application to general object recognition. In [19], a probabilistic framework was proposed to align the class distributions of two domains for robust hyperspectral image classification. Manifold alignment (MA) [20] is also a powerful tool for modeling this kind of issue. Inspired by MA, Tuia *et al.* [7] aligned multiview remote sensing images on manifolds by fully allowing for the spectral variabilities between the different angle imageries, yielding a significant improvement of classification performance.

[1]In contrast to multimodal learning (bimodality, for example), cross-modal learning trains on single modality and tests on bimodality, or *vice versa* (train on bimodality and test on single modality).

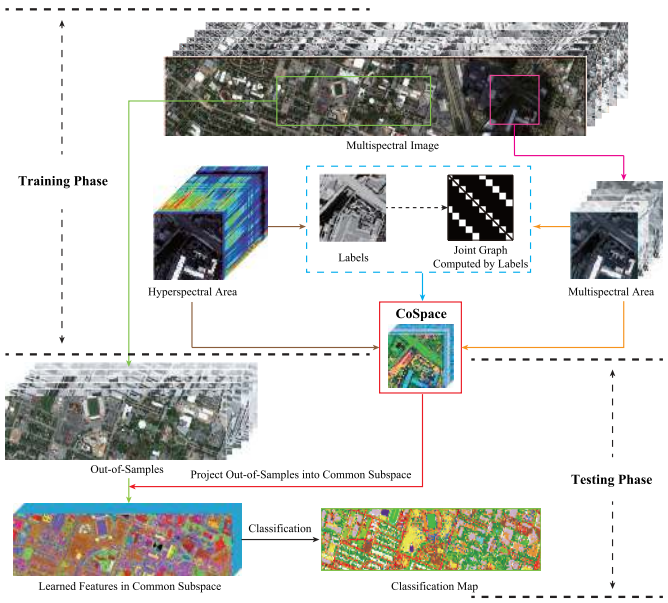


Fig. 2. Holistic workflow of the proposed CoSpace.

It should be noted that these methods mentioned earlier only consider the differences of a unimodality between the source and target domains at the level of original features, but they fail to investigate the transferability of multimodality since the different modalities usually hold the different feature dimensions. Although these approaches can build connections between features or instances, a poorly connected relationship between the learned common subspace and label information is still hindering the low-dimensional feature representation from being more discriminative.

We propose a cross-modality feature learning framework, called common subspace learning (CoSpace), that learns the shared feature representation (common subspace) from partial HS-MS correspondences. Extensive experiments are conducted on simulated MS and partially overlapped real HS data based on two airborne HS data sets: the University of Houston and Chikusei data sets. MS data are generated from HS data by using the spectral response functions (SRFs) of Sentinel-2. We relabel the training and testing classes on the data sets to meet the problem setting of cross-modal feature learning and further to make them more challenging (see Section III for details). Our contributions can be specifically unfolded as follows.

1) We propose a novel CoSpace approach by jointly considering the subspace learning and classification in order to effectively bridge the learned features and label information, aiming at addressing the HS-MS cross-modal feature learning issue.

2) By locally aligning HS-MS data on the low-dimensional manifolds where the features of HS and MS share the same dimension, CoSpace linearly learns a latent shared subspace from HS-MS correspondences, where samples are expected to be classified better. Because of the subspace learned in a linear way, the out-of-samples data can be simply and smoothly embedded.
3) An optimization algorithm based on the alternating direction method of multipliers (ADMMs) is designed to solve the proposed model.

The remainder of this paper is organized as follows. In Section II, we first clarify our motivation and then propose the methodology of the CoSpace model, finally, elaborate on the corresponding ADMM-based optimization algorithm. Section III presents the experimental results and analysis on two different HS-MS data sets both qualitatively and quantitatively. Finally, some conclusions are drawn in Section IV.

II. CoSpace: Common Subspace Learning

To take the benefit of HS imagery covering only a limited part of the MS imagery and, subsequently, improve the classification results of the entire area covered by the MS imagery, our idea is to learn an HS-MS common subspace, in which the data from one domain can be adaptively transferred to another domain.

Our solution to the problem is to learn an HS-MS common subspace, in which the data from one domain can be adaptively transferred to another domain.

Fig. 2 shows the holistic diagram of the proposed CoSpace method.

A. Problem Formulation

Let $\mathbf{X}_M \in \mathbb{R}^{d_M \times N}$ and $\mathbf{X}_H \in \mathbb{R}^{d_H \times N}$ be the observed MS image with d_M bands by N pixels and the HS image with d_H bands by N pixels, respectively. $\mathbf{Y} \in \mathbb{R}^{L \times N}$ is the label matrix represented by one-hot encoding. $\Theta_M \in \mathbb{R}^{d \times d_M}$ ($\Theta_H \in \mathbb{R}^{d \times d_H}$) is denoted as the projection matrix for connecting the MS (HS) data and the latent subspace. The variable $\mathbf{P} \in \mathbb{R}^{L \times d}$ is the weighted matrix specified by bridging the latent subspace and label information. Accordingly, $\widetilde{\mathbf{Y}} = [\mathbf{Y}, \mathbf{Y}] \in \mathbb{R}^{L \times 2N}$ can be modeled as follows.

The CoSpace can be modeled as follows:

$$\widetilde{\mathbf{Y}} = \mathbf{P}\Theta\widetilde{\mathbf{X}} + \mathbf{E} \tag{1}$$

where $\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_M & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_H \end{bmatrix} \in \mathbb{R}^{(d_M+d_H)\times 2N}$, and $\Theta = [\Theta_M, \Theta_H] \in \mathbb{R}^{d \times (d_M+d_H)}$. $\mathbf{E} \in \mathbb{R}^{L \times 2N}$ is the corresponding residual matrix containing the additive noise and other errors.

Since (1) is a typically ill-posed problem because of more degrees of flexibility involved (e.g., latent subspace estimation), several assumptions (or prior knowledge) should be introduced into CoSpace using regularization technique. Followed by a popular joint learning framework proposed in [21], we formulate the CoSpace as the following constrained optimization problem:

$$\min_{\mathbf{P},\Theta} \left\{ \begin{array}{c} \frac{1}{2}\|\widetilde{\mathbf{Y}} - \mathbf{P}\Theta\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2 + \Phi(\mathbf{P}) + \Psi(\Theta) \\ \text{s.t.} \quad \Theta\Theta^{\mathrm{T}} = \mathbf{I} \end{array} \right\}. \tag{2}$$

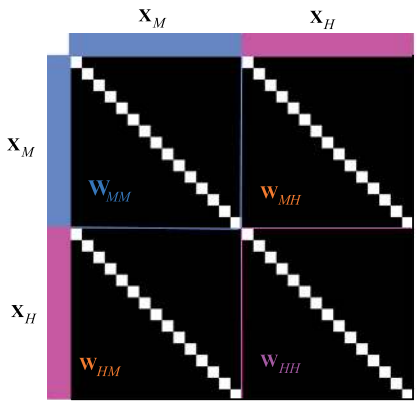


Fig. 3. Example to clarify the joint adjacency matrix.

The two regularization terms in (2) are detailed in the following.

To achieve a reliable generalization of our model, the variable $\mathbf{P}$ parameterized by α can be regularized by a Frobenius norm

$$\Phi(\mathbf{P}) = \frac{\alpha}{2}\|\mathbf{P}\|_{\mathrm{F}}^2 \tag{3}$$

and the prior knowledge with respect to Θ, resulting in a multimodal MA regularization, can be expressed with a joint graph structure as

$$\Psi(\Theta) = \frac{\beta}{2}\operatorname{tr}(\Theta\widetilde{\mathbf{X}}\mathbf{L}(\Theta\widetilde{\mathbf{X}})^{\mathrm{T}}) \tag{4}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W} \in \mathbb{R}^{2N \times 2N}$ stands for a joint Laplacian matrix, $\mathbf{W}$ that is a corresponding adjacency matrix can be directly inferred from label information in the form of the linear discriminant analysis (LDA)-like graph [22]

$$\mathbf{W}_{i,j} = \begin{cases} 1/N_k, & \text{if } \mathbf{X}_i \text{ and } \mathbf{X}_j \text{ belong to the } k\text{th class} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

and then $\mathbf{D}$ is computed by $\mathbf{D}_{ii} = \sum_{i \neq j} \mathbf{W}_{i,j}$. Fig. 3 illustrates the joint graph structure.

B. Model Optimization

Considering the nonconvexity of problem (2), an iterative alternating optimization strategy is adopted to solve the convex subproblems of each variable $\mathbf{P}$ and Θ. An implementation of CoSpace is given in Algorithm 1.

Optimization with respect to $\mathbf{P}$: This is a typical least-squares problem with the Tikhonov regularization that can be formulated as

$$\min_{\mathbf{P}} \left\{ \frac{1}{2}\|\widetilde{\mathbf{Y}} - \mathbf{P}\Theta\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2 + \frac{\alpha}{2}\|\mathbf{P}\|_{\mathrm{F}}^2 \right\} \tag{6}$$

which has a closed-form solution

$$\mathbf{P} = (\widetilde{\mathbf{Y}}\mathbf{Q}^{\mathrm{T}})(\mathbf{Q}\mathbf{Q}^{\mathrm{T}} + \alpha\mathbf{I})^{-1} \tag{7}$$

where $\mathbf{Q} = \Theta\widetilde{\mathbf{X}}$.

Algorithm 1: CoSpace

Input: $\widetilde{\mathbf{Y}}$, $\widetilde{\mathbf{X}}$, $\mathbf{L}$, and parameters α, β, *maxIter*.
Output: $\mathbf{P}$, Θ
$t = 1$, $\zeta = 1e-4$;
Initializating $\mathbf{P}$ and Θ
while *not converged* or $t >$ *maxIter* **do**
  Fix other variables to update $\mathbf{P}$ by (7)
  Fix other variables to update Θ by **Algorithm 2**
  Compute the objective function value E^{t+1} and check the convergence condition: **if** $|\frac{E^{t+1}-E^t}{E^t}| < \zeta$ **then**
    Stop iteration;
  else
    $t \leftarrow t+1$;
  end
end

Algorithm 2: Solving the Subproblem for Θ

Input: $\widetilde{\mathbf{Y}}$, $\mathbf{P}$, $\mathbf{J}$, $\widetilde{\mathbf{X}}$, $\mathbf{L}$, β, *maxIter*.
Output: Θ.
Initialization: $\Theta = \mathbf{0}$, $\mathbf{G} = \mathbf{0}$, $\Lambda_1 = \mathbf{0}$, $\Lambda_2 = \mathbf{0}$, $\mu = 10^{-3}$, $\mu_{\max} = 10^6$, $\rho = 1.5$, $\varepsilon = 10^{-6}$, $t = 1$.
while *not converged* or $t >$ *maxIter* **do**
  Fix other variables to update $\mathbf{J}$ by

$$\mathbf{J} = (\mathbf{P}^{\mathrm{T}}\mathbf{P} + \mu\mathbf{I})^{-1}(\mathbf{P}^{\mathrm{T}}\widetilde{\mathbf{Y}} + \mu\Theta\widetilde{\mathbf{X}} - \Lambda_1).$$

  Fix other variables to update Θ by

$$\Theta = (\mu\mathbf{J}\widetilde{\mathbf{X}}^{\mathrm{T}} + \Lambda_1\widetilde{\mathbf{X}}^{\mathrm{T}} + \mu\mathbf{G} + \Lambda_2) \times(\mu\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{T}} + \mu\mathbf{I} + \beta\widetilde{\mathbf{X}}\mathbf{L}\widetilde{\mathbf{X}}^{\mathrm{T}})^{-1}.$$

  Fix other variables to update $\mathbf{G}$ by

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \mathrm{svd}(\Theta - \Lambda_2/\mu), \quad \mathbf{G} = \mathbf{U}\mathbf{I}_{n\times m}\mathbf{V}.$$

  Update Lagrange multipliers by

$$\Lambda_1 \leftarrow \Lambda_1 + \mu(\mathbf{J} - \Theta\widetilde{\mathbf{X}}), \quad \Lambda_2 \leftarrow \Lambda_2 + \mu(\mathbf{G} - \Theta).$$

  Update penalty parameter by

$$\mu = \min(\rho\mu, \mu_{\max}).$$

  Check the convergence conditions: **if** $\|\mathbf{J} - \Theta\widetilde{\mathbf{X}}\|_{\mathrm{F}} < \varepsilon$ *and* $\|\mathbf{G} - \Theta\|_{\mathrm{F}} < \varepsilon$ **then**
    Stop iteration;
  else
    $t \leftarrow t+1$;
  end
end

Optimization with respect to Θ: The optimization problem for Θ can be formulated as

$$\min_{\Theta}\left\{\begin{array}{c}\frac{1}{2}\|\widetilde{\mathbf{Y}} - \mathbf{P}\Theta\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2 + \frac{\beta}{2}\mathrm{tr}(\Theta\widetilde{\mathbf{X}}\mathbf{L}(\Theta\widetilde{\mathbf{X}})^{\mathrm{T}}) \\ \text{s.t.} \quad \Theta\Theta^{\mathrm{T}} = \mathbf{I}\end{array}\right\}. \tag{8}$$

In order to solve (8) effectively with ADMM, we consider an equivalent form by introducing auxiliary variables $\mathbf{J}$ and $\mathbf{G}$ to replace $\Theta\widetilde{\mathbf{X}}$ and Θ, respectively,

$$\min_{\Theta,\mathbf{J},\mathbf{G}}\left\{\begin{array}{c}\frac{1}{2}\|\widetilde{\mathbf{Y}} - \mathbf{P}\mathbf{J}\|_{\mathrm{F}}^2 + \frac{\beta}{2}\mathrm{tr}(\Theta\widetilde{\mathbf{X}}\mathbf{L}(\Theta\widetilde{\mathbf{X}})^{\mathrm{T}}) \\ \text{s.t.} \quad \mathbf{J} = \Theta\widetilde{\mathbf{X}}, \ \mathbf{G} = \Theta, \ \mathbf{G}\mathbf{G}^{\mathrm{T}} = \mathbf{I}\end{array}\right\}. \tag{9}$$

The augmented Lagrangian version of (9) is

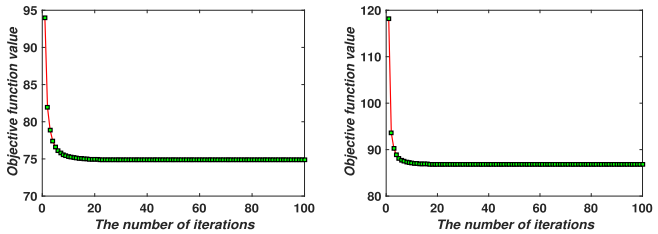
$$\begin{aligned}&\mathscr{L}_C(\Theta, \mathbf{J}, \mathbf{G}, \Lambda_1, \Lambda_2)\\ &\quad= \frac{1}{2}\|\widetilde{\mathbf{Y}} - \mathbf{P}\mathbf{J}\|_{\mathrm{F}}^2 + \frac{\beta}{2}\mathrm{tr}(\Theta\widetilde{\mathbf{X}}\mathbf{L}(\Theta\widetilde{\mathbf{X}})^{\mathrm{T}}) + \Lambda_1^{\mathrm{T}}(\mathbf{J} - \Theta\widetilde{\mathbf{X}})\\ &\qquad+ \Lambda_2^{\mathrm{T}}(\mathbf{G} - \Theta) + \frac{\mu}{2}\|\mathbf{J} - \Theta\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2 + \frac{\mu}{2}\|\mathbf{G} - \Theta\|_{\mathrm{F}}^2\\ &\qquad\text{s.t.} \quad \mathbf{G}\mathbf{G}^{\mathrm{T}} = \mathbf{I}\end{aligned} \tag{10}$$

where Λ_1 and Λ_2 are the Lagrange multipliers and μ is the penalty parameter. Algorithm 2 summarizes the specific procedures for solving the problem (9), and the solution to each subproblem is detailed in Appendix A.

Finally, we repeat these optimization procedures until a stopping criterion is satisfied.

C. Convergence Analysis

The iterative alternating strategy used in Algorithms 1 and 2 is a block coordinate descent, whose convergence is theoretically guaranteed as long as each subproblem of (2) is strictly convex, which can be exactly minimized [23]. Moreover, we experimentally display an illustration to clarify the convergence of CoSpace on both HS-MS data sets, where the objective function value is recorded in each iteration (see Fig. 4).



(a) (b)

Fig. 4. Convergence analysis of CoSpace is experimentally performed on the two HS-MS data sets. (a) University of houston HS-MS data sets. (b) Chikusei HS-MS data sets.

III. Experiments

In this section, we quantitatively and qualitatively evaluate the performance of the proposed method on two HS-MS data sets taken over the University of Houston and Chikusei. To validate the transferability of learned features by our CoSpace method, classification is explored as a potential application. Therefore, three different classifiers, namely, the nearest neighbor (NN) based on the Euclidean distance, linear support vector machines (LSVMs), and canonical correlation forest (CCF) [24], are selected for this task. As a variant of random forest [25], CCF has shown its effectiveness in various tasks [26]–[28] because of supervised feature extraction via canonical correlation analysis when constructing each decision tree. Furthermore, we compare the proposed method (CoSpace)

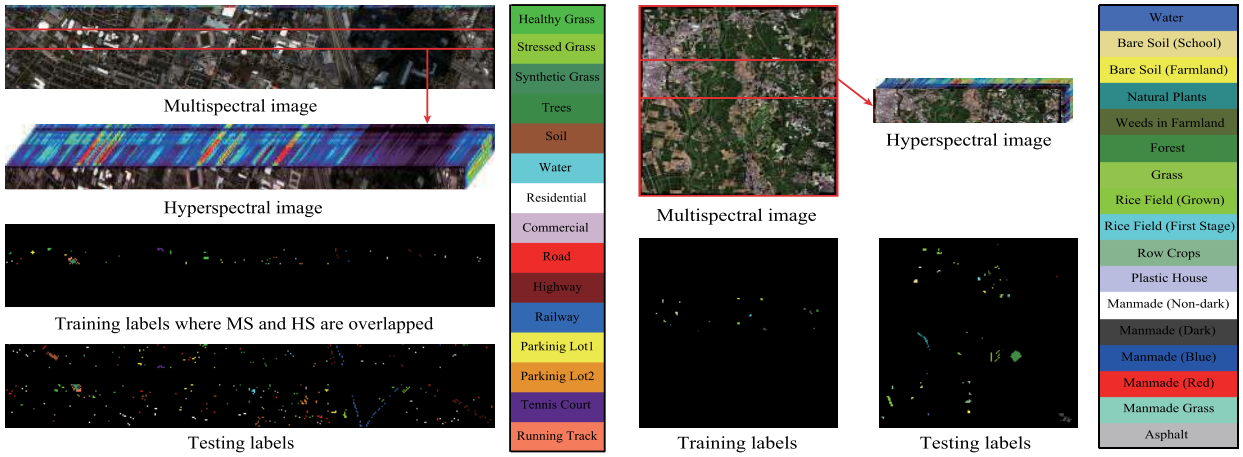


Fig. 5. MS image and its corresponding hyperspectral image that partially covers the same area, as well as training and testing labels, for (a) University of Houston HS-MS data set and (b) Chikusei HS-MS data set, respectively.

TABLE I
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE UNIVERSITY OF HOUSTON MS-HS DATA SET

Class No.	Class Name	Training	Testing
1	Healthy Grass	537	699
2	Stressed Grass	61	1154
3	Synthetic Grass	340	357
4	Tree	209	1035
5	Soil	74	1168
6	Water	22	303
7	Residential	52	1203
8	Commercial	320	924
9	Road	76	1149
10	Highway	279	948
11	Railway	33	1185
12	Parking Lot1	329	904
13	Parking Lot2	20	449
14	Tennis Court	266	162
15	Running Track	279	381
	Total	2897	12021

TABLE II
NUMBER OF TRAINING AND TESTING SAMPLES FOR THE CHIKUSEI MS-HS DATA SET

Class No.	Class Name	Training	Testing
1	Water	301	858
2	Bare Soil (School)	992	1867
3	Bare Soil (Farmland)	455	4397
4	Natural Plants	150	4272
5	Weeds in Farmland	928	1108
6	Forest	486	11904
7	Grass	989	5526
8	Rice Field (Grown)	813	8816
9	Rice Field (First Stage)	667	1268
10	Row Crops	377	5961
11	Plastic House	165	475
12	Manmade (Non-dark)	170	568
13	Manmade (Dark)	1291	6373
14	Manmade (Blue)	111	431
15	Manmade (Red)	35	187
16	Manmade Grass	21	1019
17	Asphalt	384	417
	Total	8335	55447

with several classical approaches, which are suitable for the cross-modal feature learning task, including principle component analysis based on joint dimensionality reduction (P-JDR for short) [29], locality preserving projection (LPP) based on unsupervised MA (L-USMA for short) [30], and LPP-based supervised MA (L-SMA) [31] as well as the original MS (baseline). Tables I and II list the number of training and test samples on two used data sets.

A. University of Houston HS-MS Data Sets

1) Data Description: The HS data were acquired by the ITRES CASI-1500 sensor over an urban area around the campus of the University of Houston, Houston, TX, USA, which was provided in the 2013 IEEE GRSS data fusion contest [32]. The image consists of 349 × 1905 pixels with 144 spectral bands in the wavelength from 364 to 1046 nm with spectral resolution of 10 nm at a ground sampling distance of 2.5 m. Spectral simulation is performed to generate the MS image by degrading the full HS image in the spectral domain using the MS (SRFs of Sentinel-2 as filters (for more details refer to [5]). Following this, the MS data with dimensions of $349\times1905\times10$ are generated. The MS image and the corresponding partial HS image over the University of Houston scene are shown in Fig. 5(a).

2) Experimental Setup: Initially, we redistribute the training and testing samples, as shown in Fig. 5(a) and, more specifically, listed in Table I, to meet our problem setting that there is a large amount of the MS data (complete low-quality data) together with a limited amount of the HS data (incomplete high-quality data).

For the performance assessment of the algorithms, we adopt three criteria to quantify experiential results as follows.

1) *Overall Accuracy* (OA): This index is defined by the ratio between the number of MS samples that are correctly classified and the number of corresponding test samples.

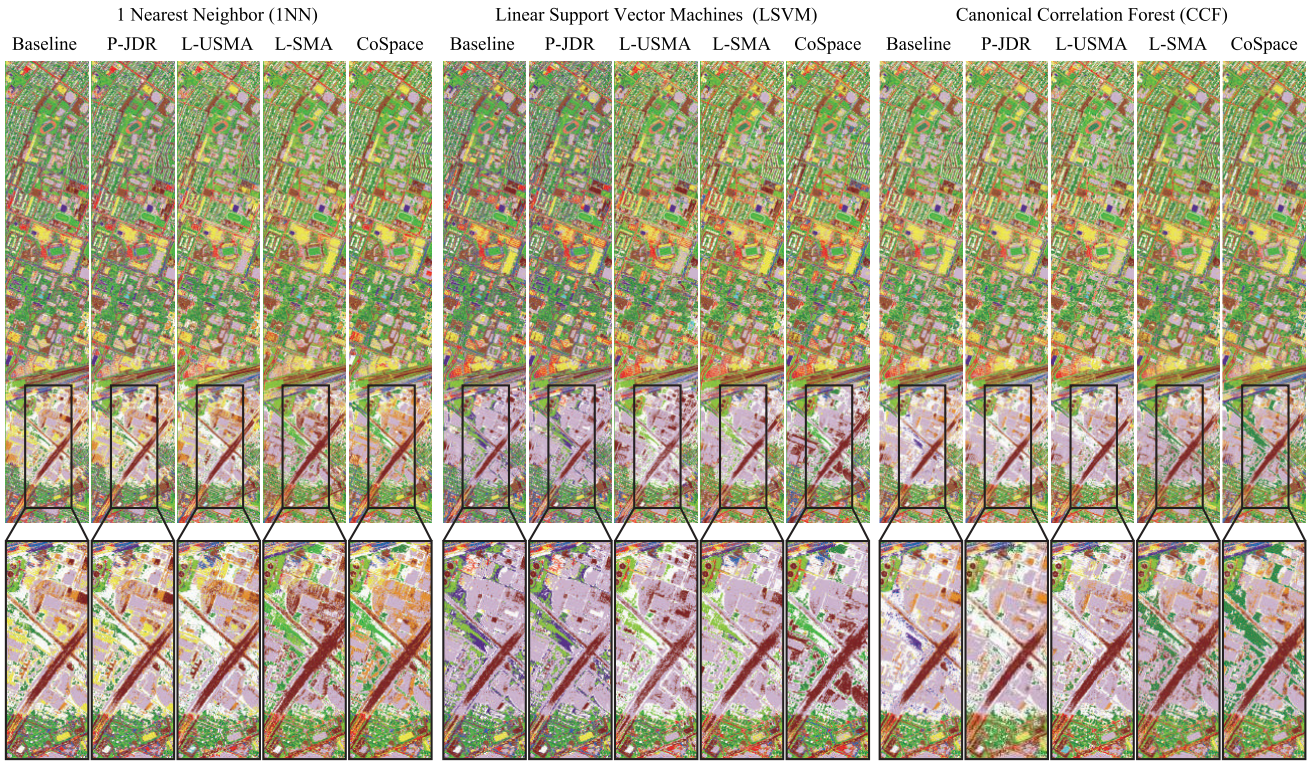


Fig. 6. Classification maps and corresponding highlighted subareas of the different algorithms obtained using three kinds of classifiers on the University of Houston data set.

2) *Average Accuracy* (AA): We collect the classification accuracy of each class and average them to achieve an AA-based evaluation.
3) *Kappa Coefficient* (κ): It statistically measures the agreement between the final classification map and the ground-truth map. Generally speaking, κ is more robust and convincing than a simple percent-based agreement calculation (e.g., OA and AA), since the agreement occurring by chance is fully considered.

Furthermore, we experimentally maximize the performance of the different algorithms by tuning their parameters, such as dimension (d), regularization parameters (α, β, γ), and so on, using ten-fold cross-validation on training data. For the dimension (d) which is a common parameter for all algorithms, they can be selected ranging from 10 to 50 at an interval of 10. For the number of NNs (k) and the standard deviation of Gaussian kernel function (σ) in L-USMA, we select them in the range of $\{10, 20, \ldots, 50\}$ and $\{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$, respectively, and two regularization parameters (α, β) in CoSpace are both chosen from $\{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$.

3) Results and Analysis: Fig. 6 shows the classification maps of compared algorithms using three different classifiers, while Table III details the quantitative assessment results under the optimal parameters determined by cross-validation.

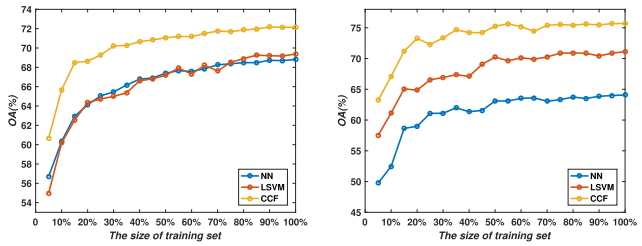
Overall, after absorbing partial HS information, those ASFL approaches are prone to obtain a better classification result, compared to the baseline (only MS data). P-JDR steadily outperforms the baseline, especially using 1NN and LSVM classifiers, although its classification accuracy using CCF is slightly lower than that of baseline. By embedding local topological structure of data, L-USMA performs better than baseline, and even P-JDR, showing stable results for three kinds of classifiers. With a more discriminative supervised information, L-SMA obtains more competitive results by locally constructing LDA-like graph, whose performance is basically superior to that of the baseline, P-JDR, and L-USMA. Unlike L-SMA that only aligns different modalities on a common subspace, the proposed CoSpace learns a latent subspace by aligning different modalities and also bridges the learned subspace with label information, achieving the best classification accuracy. Compared to baseline, P-JDR, L-USMA, and L-SMA, CoSpace increases the OAs of 7.12%, 2.45%, 2.49%, and 3.78%, respectively, with 1NN classifier, and 7.26%, 5.07%, 3.84%, and 1.37%, respectively, with LSVM classifier, as well as 3.96%, 5.04%, 3.75%, and 2.58%, respectively, with CCF classifier. Likewise, there are similar trends for the other indices of AA and κ, which indicate that CoSpace tends to learn semantically meaningful features.

We can also observe from Fig. 5(a) and Table I that the training samples collected in a very limited area badly results in the data unbalance between different classes. For instance, the number of training samples in *Health Grass* is dozens of times as much as that in *Water*, *Railway*, *Residential*, *Commercial*, and *Parking Lot2*. This might make the classifier impossible to be trained effectively, since more attentions are paid on those classes with large-size samples, and, contrariwise, the small-scale classes play relatively less and even nothing.

TABLE III

QUANTITATIVE PERFORMANCE COMPARISON WITH THE DIFFERENT ALGORITHMS ON THE UNIVERSITY OF HOUSTON DATA. THE BEST ONE IS SHOWN IN BOLD

Algorithm	Baseline (%)			P-JDR (%)			L-USMA (%)			L-SMA (%)			CoSpace (%)		
Parameter	l			$d = 30$			$(k, \sigma, d) = (10, 1, 20)$			$d = 30$			$(\alpha, \beta, d) = (0.01, 0.01, 30)$		
Classifier	1NN	LSVM	CCF	1NN	LSVM	CCF	1NN	LSVM	CCF	1NN	LSVM	CCF	1NN	LSVM	CCF
OA	61.70	62.12	68.21	66.37	64.31	67.13	66.33	65.54	68.41	65.04	68.01	69.59	68.82	69.38	**72.17**
AA	65.57	65.97	70.47	69.23	67.50	69.64	69.37	68.81	70.84	68.15	70.50	71.02	71.29	71.69	**73.56**
κ	0.5842	0.5889	0.6543	0.6345	0.6118	0.6430	0.6342	0.6251	0.6565	0.6202	0.6520	0.6695	0.6613	0.6672	**0.6975**
Class1	69.10	76.39	67.95	73.39	71.82	70.24	80.40	78.68	67.67	81.69	75.25	68.53	**88.56**	75.54	69.96
Class2	79.20	80.59	78.08	81.20	81.72	72.53	78.94	79.90	75.04	90.99	**97.57**	77.90	73.48	73.74	77.99
Class3	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Class4	85.22	85.51	92.27	87.63	91.21	93.24	94.30	96.23	92.85	88.60	94.78	**98.74**	93.62	**98.74**	98.26
Class5	98.89	99.06	99.40	98.72	98.72	99.32	**99.49**	99.40	98.89	99.40	98.97	99.14	98.97	99.40	99.40
Class6	86.14	86.14	86.14	86.47	86.14	76.57	86.14	**86.47**	**86.47**	**86.47**	**86.47**	70.96	85.15	85.48	85.15
Class7	36.49	50.62	63.76	59.6	53.78	51.04	49.79	50.21	63.18	58.19	72.32	77.14	69.99	73.98	**80.05**
Class8	50.76	56.49	56.06	59.42	59.42	59.09	54.76	**66.23**	56.82	55.74	62.01	62.23	58.77	63.53	62.01
Class9	60.92	56.22	70.58	64.14	63.01	**72.32**	63.45	65.19	69.10	47.17	49.96	61.27	64.40	59.79	64.93
Class10	40.93	45.36	45.25	44.41	49.16	45.36	44.09	53.90	47.47	49.47	58.12	52.32	45.15	**64.14**	57.70
Class11	39.16	27.43	43.88	36.03	35.53	39.41	45.91	29.79	43.71	35.36	28.86	36.46	44.14	36.54	**47.26**
Class12	41.37	31.64	56.08	51.55	25.66	**65.82**	45.8	29.76	62.06	65.04	35.84	62.5	50.22	46.79	62.72
Class13	0.45	0.00	0.67	0.67	0.00	2.45	0.22	0.00	1.11	**4.68**	0.00	0.00	0.89	0.00	0.45
Class14	97.53	97.53	98.77	98.15	98.77	99.38	99.38	98.77	**100.00**	97.53	**100.00**	**100.00**	98.15	**100.00**	99.38
Class15	97.38	96.59	**98.16**	97.11	97.64	97.9	97.9	97.64	**98.16**	97.9	97.38	**98.16**	97.9	97.64	**98.16**



(a) (b)

Fig. 7. Sensitivity analysis to the training set size using three different classifiers on the two MS-HS data sets. (a) Houston HS-MS data sets. (b) Chikusei HS-MS data sets.

A feasible solution to the problem is to enhance data representative ability by jointly feature learning from multimodalities. For the performance evaluation of classifying those small-scale classes, e.g., *Residential*, *Commercial*, and *Railway*, a direct evidence has been shown in Table III that those ASFL-based approaches (e.g., P-JDR, L-USMA, and L-SMA as well as CoSpace) obviously perform better on these small-scale samples than directly using original MS data (baseline). As expected, CoSpace dramatically outperforms the others, particularly on *Residential* and *Railway*. There is no denying, however, that CoSpace is superior to other algorithms to a larger extent, although it fails to effectively identify *Parking Lot2* as same with others.

To visually highlight the classification differences for the different methods, we enlarge the classification maps of a subarea overshadowed by the cloud, as shown in Fig. 6 where we can see that the methods with considering the hyperspectral information are able to generate the more discriminative features than the baseline, while the proposed CoSpace yields a better performance in identifying the materials in the shadow area, particularly for vegetation (e.g., *Grass*), *Residential*, and *Commercial* that are easily misclassified by the traditional methods.

4) Sensitivity Analysis to the Training Set Size: As the performance of the CoSpace largely depends on the number of training samples, it is, therefore, indispensable to investigate the sensitivity of the training set size. In detail, we conduct the classification using the CoSpace by fixing the test set and setting a series of new training sets randomly selected from the original training set with the different percentages ranging from 5% to 100% at a 5% interval. As can be seen in Fig. 7(a), there is a similar trend in OAs using different classifiers, that is, the classification accuracy improves with the training set size, faster in the early, and later basically stabilized.

B. Chikusei HS-MS Data Sets

1) Data Description: The Headwall's hyperspectral visible and near-infrared series C (VNIR-C) imaging sensor acquired the airborne HS data set over the agricultural and urban areas of Chikusei, Ibaraki, Japan, in 2014. This VNIR-C sensor collected 128 bands covering the wavelength range from 363 to 1018 nm with spectral resolution of 10 nm, and the scene consists of 2517 × 2335 pixels at ground sample distance of 2.5 m. The data set was made available to the scientific community recently, and more details regarding the data acquisition and processing can be found in [33]. Similarly, the MS image with the size of $2517 \times 2335 \times 10$ was simulated by spectrally down-sampling the full HS image using the known SRFs of Sentinel-2. The generated MS image and the partial HS image over the Chikusei scene are shown in Fig. 5(b).

2) Experimental Setup: Fig. 5(b) shows the latest training and testing labeling of the Chikusei data set, which is quantified in Table II. Three indices: OA, AA, and κ introduced earlier are calculated to quantitatively assess the classification performance. Similar to the case of the University of Houston data set, the parameters for those given algorithms are determined by the tenfold cross-validation on the training samples and the same range setting with those used for the University of Houston data set is also conducted to the Chikusei data set.

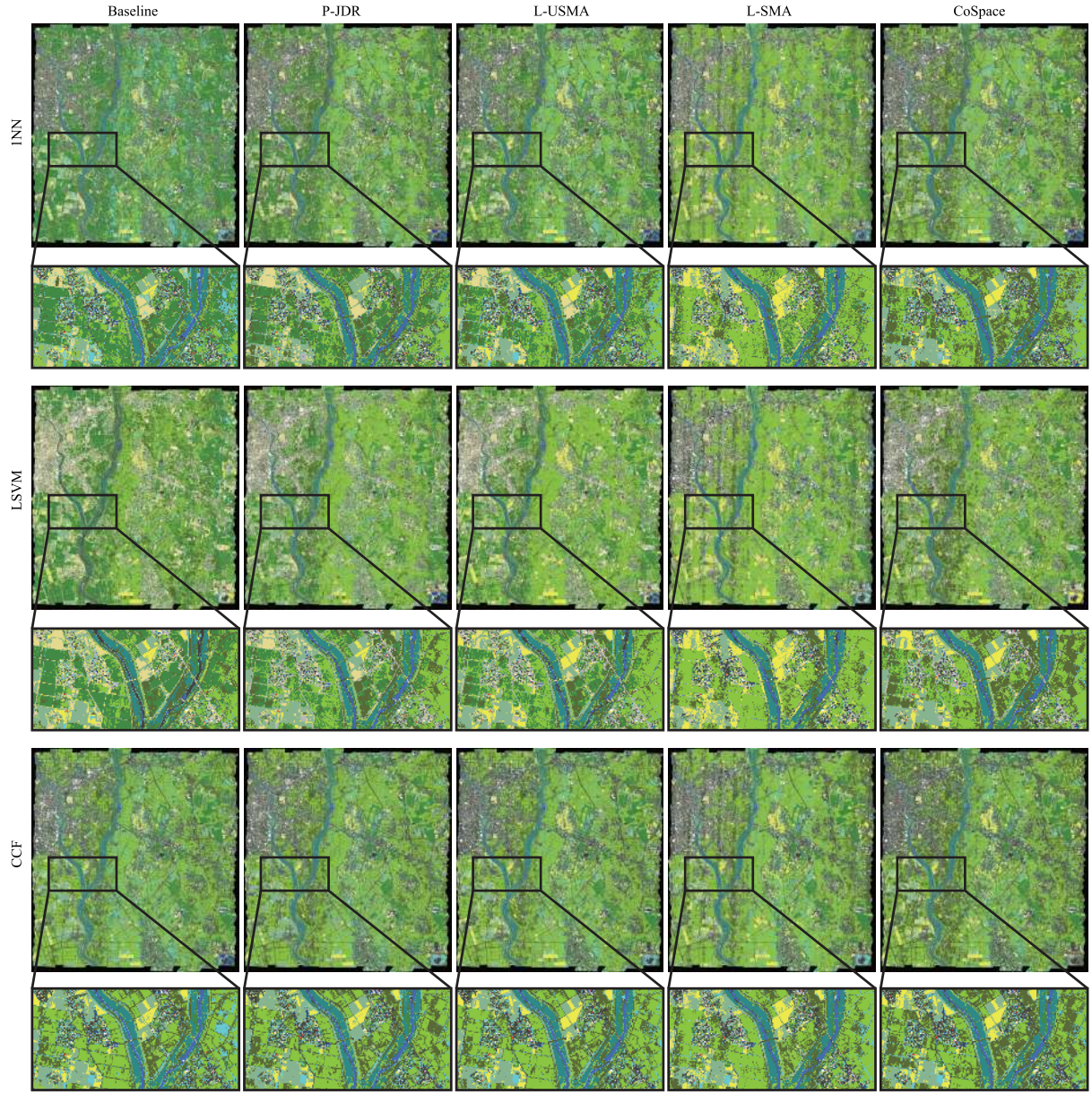


Fig. 8. Classification maps and corresponding highlighted subareas of the different algorithms obtained using three kinds of classifiers on the Chikusei data set.

3) Results and Analysis: Similar to the University of Houston scene, we evaluate the performance for the Chikusei data both quantitatively and visually. Three classification indices with optimal parameters for different algorithms are summarized in Table IV. For visual comparison, we give the corresponding classification maps in the full scene with those comparative algorithms under the different classifiers, as shown in Fig. 8.

As the classes in the Chikusei scene are more challenging classes and the distribution of training samples is inhomogeneous, directly using original MS data as input fails to identify certain materials, such as *Forest*, *Man-made (Dark)*, and *Man-made (Grass)*, yielding a poor performance in OA, AA, and κ. Especially while using 1NN classifier, P-JDR and L-USMA, which belong to the unsupervised feature learning method, observably exceed baseline in classification accuracy by 4.46% and 5.49%, respectively. For LSVM and CCF classifiers, a similar trend is also demonstrated in Table IV. Because of the limited training samples and their distribution unbalance, the subspace projection learned by L-SMA easily traps into over-fitting, despite only having a weak performance improvement compared to these previously compared algorithms. By jointly performing subspace learning and classification, CoSpace not only aligns the different modalities in a latent common subspace but also connects the subspace with label information formulated by training data. As a result, CoSpace obtains a higher classification accuracy than other algorithms, as listed in Table IV. This might attribute to

TABLE IV
QUANTITATIVE PERFORMANCE COMPARISON WITH THE DIFFERENT ALGORITHMS ON THE CHIKUSEI DATA. THE BEST ONE IS SHOWN IN BOLD

Algorithm	Baseline (%)			P-JDR (%)			L-USMA (%)			L-SMA (%)			CoSpace (%)		
Parameter	l			$d = 30$			$(k, \sigma, d) = (10, 1, 30)$			$d = 20$			$(\alpha, \beta, d) = (0.1, 0.01, 30)$		
Classifier	1NN	LSVM	CCF	1NN	LSVM	CCF	1NN	LSVM	CCF	1NN	LSVM	CCF	1NN	LSVM	CCF
OA	55.99	60.20	71.11	60.45	68.19	71.87	61.48	67.31	72.33	62.44	67.90	71.53	64.07	71.12	**75.69**
AA	62.39	69.42	70.40	64.95	71.71	71.02	64.95	70.06	71.49	64.52	70.79	66.47	68.05	**73.96**	71.46
κ	0.5084	0.5523	0.6761	0.5614	0.6414	0.6834	0.5715	0.6318	0.6893	0.5812	0.6391	0.6802	0.5995	0.6746	**0.7260**
Class1	79.49	78.21	80.54	97.2	98.25	81.93	98.48	**99.18**	81.00	80.89	98.72	82.52	80.19	92.54	79.25
Class2	**95.02**	94.43	82.7	93.2	93.95	94.86	93.25	93.68	93.95	92.98	93.20	92.50	**95.02**	93.47	94.91
Class3	11.37	23.54	50.06	14.40	16.03	24.13	31.07	39.37	56.74	61.81	62.57	55.31	80.10	**80.40**	77.71
Class4	91.64	92.13	92.56	88.65	90.87	93.38	89.49	92.21	94.90	86.54	90.57	91.53	93.47	90.59	**96.23**
Class5	87.00	**97.65**	94.68	67.78	70.94	76.26	63.90	34.57	79.87	25.36	28.43	16.06	74.28	83.94	66.52
Class6	65.36	62.01	81.48	55.18	75.11	**87.77**	50.90	67.15	79.42	48.29	62.52	78.91	53.49	63.61	79.02
Class7	99.26	99.67	99.93	91.95	98.43	**99.98**	91.02	97.00	99.71	95.46	96.87	97.79	88.74	97.74	99.75
Class8	43.42	57.11	93.40	85.32	95.58	97.95	91.64	96.32	**99.29**	93.93	95.59	93.49	65.61	95.05	92.72
Class9	99.92	**100.00**	**100.00**	98.34	98.66	99.53	98.82	99.45	99.92	99.21	99.53	99.13	**100.00**	98.66	99.76
Class10	16.96	**24.81**	19.56	21.36	23.75	22.24	19.78	20.13	17.16	21.15	21.39	15.48	22.68	22.35	18.00
Class11	2.11	0.00	2.11	0.63	1.05	**6.53**	0.00	0.00	3.37	0.00	0.00	0.00	0.00	0.00	0.00
Class12	87.85	90.32	88.91	86.44	88.20	89.96	81.51	88.38	87.32	84.43	90.14	85.92	86.62	**90.32**	80.46
Class13	33.08	33.11	33.09	33.08	33.11	33.11	32.36	33.01	33.11	25.47	32.61	56.25	33.11	33.11	**67.90**
Class14	67.75	94.20	85.38	59.40	58.24	**99.77**	58.70	59.40	89.56	71.93	72.85	59.40	59.40	59.40	52.44
Class15	97.33	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	96.26	96.79	**100.00**	94.12	93.58	**100.00**	**100.00**	93.58	97.86
Class16	66.24	74.88	88.62	95.39	95.58	99.71	99.80	98.43	**100.00**	99.61	99.71	99.51	93.13	97.84	**100.00**
Class17	16.79	58.03	3.84	15.83	**81.29**	0.24	7.19	76.26	0.00	15.59	65.23	7.91	30.94	64.75	0.00

the learned common subspace, since the features projected in the subspace can absorb various properties from different modalities.

Similarly, we also make a visual comparison by giving a salient region in which the CoSpace's superiority in classifying complex and similar land-cover classes is further shown as detailed in Fig. 8. Compared to other alignment-based methods, CoSpace is capable of better transferring HS information into MS data by means of joint subspace learning and classification, yielding a more discriminative low-dimensional embedding. The learned features can recognize those classes of holding very similar features in MS data, such as *Bare Soil (Farmland)* and *Row Crops*, *Weeds in Farmland*, and *Rice Field (Grown)*, more effectively. As shown in Fig. 8, CoSpace performs more reasonable and competitive classification results, that is, on the one hand, the *Weeds in Farmland* and *Rice Field (Grown)* are most likely to be coexisted in a scene; on the other hand, the *Bare Soil (Farmland)* and *Row Crops* are separated more correctly. This can be explained by a powerful transferability of HS information in the proposed CoSpace.

4) Sensitivity Analysis to the Training Set Size: Similar to the MS-HS Houston data sets, we apply the same investigating strategy and observe the trend of classification performance using CoSpace with different sizes of training sets on the MS-HS Chikusei data sets in Fig. 7(b). There is a very substantial change in classification accuracy with the increase of the training set size ranging from 5% to 40% of total training samples, while the performance tends to be stable after the training set size is over 50%.

IV. CONCLUSION

The tradeoff between MS and HS imaging in terms of observation ranges and spectral resolution motivates us to ponder whether HS data partially overlapping MS data can contribute to improving the classification performance of the whole MS imagery. For this purpose, we proposed CoSpace to achieve the property transferring in the different domains by learning a latent common subspace. Moreover, an effective joint strategy that simultaneously considers subspace learning and classification is embedded into the proposed method to tightly bridge the gap between the learned subspace and label information, leading to a more discriminative feature representation. The superior classification performance using CoSpace is demonstrated on two different data sets, compared to using other state-of-art methods.

We performed transfer learning on homogeneous data sets in the considered MS-HS case in the sense that both data sources are optical images covering similar spectral ranges and thus the HS information can be transferred into the MS one linearly. The CoSpace's ability in handling heterogeneous data sources remains limited due to its linearized modeling. In the future work, we will develop a more general system by integrating some powerful and emerging nonlinear tools (e.g., deep learning) into our framework.

In addition, we just assumed to share the same land-cover classes across MS and HS images in this paper. In reality, the number of land-cover classes in the large-scale MS scene might be usually more than the one in the overlapped area of MS and HS images. This naturally motivates us to generalize our model in the future work.

APPENDIX
SOLUTION TO PROBLEM (8) WITH RESPECT TO Θ

The solution to problem (8) can be transferred to equivalently solve the problem (10) with ADMM. Considering the fact that the object function in (10) is not convex with respect to all variables simultaneously, but it is a convex problem regarding the separate variable when other variables are fixed, therefore, we successively minimize $\mathscr{L}_C$ with respect to $\Theta, \mathbf{J}, \mathbf{G}, \Lambda_1, \Lambda_2$ as follows.

Optimization with respect to Θ*:* The optimization problem for Θ can be written as

$$\min_{\Theta}\left\{\begin{array}{c}\frac{\beta}{2}\operatorname{tr}(\Theta\widetilde{\mathbf{X}}\mathbf{L}(\Theta\widetilde{\mathbf{X}})^{\mathrm{T}})+\Lambda_1^{\mathrm{T}}(\mathbf{J}-\Theta\widetilde{\mathbf{X}})\\ +\Lambda_2^{\mathrm{T}}(\mathbf{G}-\Theta)+\frac{\mu}{2}\|\mathbf{J}-\Theta\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2+\frac{\mu}{2}\|\mathbf{G}-\Theta\|_{\mathrm{F}}^2\}\end{array}\right\} \tag{11}$$

which has a closed-form solution

$$\Theta\leftarrow(\mu\mathbf{J}\widetilde{\mathbf{X}}^{\mathrm{T}}+\Lambda_1\widetilde{\mathbf{X}}^{\mathrm{T}}+\mu\mathbf{G}+\Lambda_2)\times(\mu\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathrm{T}}+\mu\mathbf{I}+\beta\widetilde{\mathbf{X}}\mathbf{L}\widetilde{\mathbf{X}}^{\mathrm{T}})^{-1}. \tag{12}$$

Optimization with respect to $\mathbf{J}$*:* The variable $\mathbf{J}$ can be estimated by solving the following problem:

$$\min_{\mathbf{J}}\left\{\frac{1}{2}\|\widetilde{\mathbf{Y}}-\mathbf{P}\mathbf{J}\|_{\mathrm{F}}^2+\Lambda_1^{\mathrm{T}}(\mathbf{J}-\Theta\widetilde{\mathbf{X}})+\frac{\mu}{2}\|\mathbf{J}-\Theta\widetilde{\mathbf{X}}\|_{\mathrm{F}}^2\right\} \tag{13}$$

its analytical solution is given by

$$\mathbf{J}\leftarrow(\mathbf{P}^{\mathrm{T}}\mathbf{P}+\mu\mathbf{I})^{-1}(\mathbf{P}^{\mathrm{T}}\widetilde{\mathbf{Y}}+\mu\Theta\widetilde{\mathbf{X}}-\Lambda_1). \tag{14}$$

Optimization with respect to $\mathbf{G}$*:* For $\mathbf{G}$, the optimization problem with orthogonal constraint can be formulated as

$$\min_{\mathbf{G}}\left\{\Lambda_2^{\mathrm{T}}(\mathbf{G}-\Theta)+\frac{\mu}{2}\|\mathbf{G}-\Theta\|_{\mathrm{F}}^2\},\ \text{s.t. }\mathbf{G}\mathbf{G}^{\mathrm{T}}=\mathbf{I}\right\} \tag{15}$$

which can be effectively solved using the strategy of splitting orthogonality constraints [34] in two steps.

The first step is to perform the *singular value decomposition (SVD) factorization*

$$[\mathbf{U},\mathbf{S},\mathbf{V}]=\operatorname{svd}(\Theta-\Lambda_2/\mu). \tag{16}$$

The second step is to update $\mathbf{G}$ with satisfying orthogonal constraint

$$\mathbf{G}\leftarrow\mathbf{U}\mathbf{I}_{n\times m}\mathbf{V}. \tag{17}$$

Lagrange multipliers (Λ_1, Λ_2) *and penalty parameter* (μ) *update:* Before stepping into the next iteration, Lagrange multipliers need to be updated by

$$\Lambda_1\leftarrow\Lambda_1+\mu(\mathbf{J}-\Theta\widetilde{\mathbf{X}}),\quad \Lambda_2\leftarrow\Lambda_2+\mu(\mathbf{G}-\Theta) \tag{18}$$

and penalty parameter be updated by

$$\mu\leftarrow\ \min(\rho\mu,\mu_{\max}). \tag{19}$$

ACKNOWLEDGMENT

The authors would like to thank the Hyperspectral Image Analysis Group and the NSF Funded Center for Airborne Laser Mapping at the University of Houston for providing the CASI University of Houston data set. They would also like to thank Prof. D. Cai and Dr. C. Wang for providing MATLAB codes for LPP and MA algorithms.

REFERENCES

[1] P. Du, S. Liu, P. Gamba, K. Tan, and J. Xia, "Fusion of difference images for change detection over urban areas," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1076–1086, Aug. 2012.

[2] C. Yang, J. H. Everitt, Q. Du, B. Luo, and J. Chanussot, "Using high-resolution airborne and satellite imagery to assess crop growth and yield variability for precision agriculture," *Proc. IEEE*, vol. 101, no. 3, pp. 582–592, Mar. 2013.

[3] H. Xie *et al.*, "Dynamic monitoring of agricultural fires in China from 2010 to 2014 using MODIS and GlobeLand30 data," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 10, p. 172, 2016.

[4] F. D. Van der Meer, H. M. A. Van der Werff, and F. J. A. Van Ruitenbeek, "Potential of ESA'S Sentinel-2 for geological applications," *Remote Sens. Environ.*, vol. 148, pp. 124–133, Apr. 2014.

[5] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[6] Y. Chen, C. Li, P. Ghamisi, X. Jia, and Y. Gu, "Deep fusion of remote sensing data for accurate classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1253–1257, Aug. 2017.

[7] D. Tuia, M. Volpi, M. Trolliet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7708–7720, Dec. 2014.

[8] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[9] P. Ghamisi, B. Höfle, 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.*, vol. 10, no. 6, pp. 3011–3024, Dec. 2017.

[10] G. Iyer, J. Chanussot, and A. L. Bertozzi, "A graph-based approach for feature extraction and segmentation of multimodal images," in *Proc. ICIP*, Sep. 2017, pp. 3320–3324.

[11] D. Hong, N. Yokoya, and X. Zhu, "Local manifold learning with robust neighbors selection for hyperspectral dimensionality reduction," in *Proc. IGARSS*, 2016, pp. 40–43.

[12] D. Hong, N. Yokoya, and X. Zhu, "Learning a robust local manifold representation for hyperspectral dimensionality reduction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2960–2975, Jul. 2017.

[13] S. Liu, Q. Du, X. Tong, A. Samat, L. Bruzzone, and F. Bovolo, "Multi-scale morphological compressed change vector analysis for unsupervised multiple change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4124–4137, 2017.

[14] G. Tochon, J. Chanussot, M. D. Mura, and A. L. Bertozzi, "Object tracking by hierarchical decomposition of hyperspectral video sequences: Application to chemical gas plume tracking," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4567–4585, May 2017.

[15] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "Learning a low-coherence dictionary to address spectral variability for hyperspectral unmixing," in *Proc. ICIP*, Sep. 2017, pp. 1–5.

[16] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Ng, "Multimodal deep learning," in *Proc. IMLS*, Jul. 2011, pp. 689–696.

[17] G. Matasci, M. Volpi, D. Tuia, and M. Kanevski, "Transfer component analysis for domain adaptation in image classification," *Proc. SPIE*, vol. 8180, p. 81800F, Oct. 2011.

[18] B. Kulis, K. Saenko, and T. Darrell, "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," in *Proc. CVPR*, 2011, pp. 1785–1792.

[19] X. Zhou and S. Prasad, "Domain adaptation for robust classification of disparate hyperspectral images," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 822–836, Dec. 2017.

[20] C. Wang, P. Krafft, and S. Mahadevan, *Chapter of Manifold Learning: Theory and Applications-Manifold Alignment*. New York, NY, USA: CSC Press, 2011.

[21] S. Ji and J. Ye, "Linear dimensionality reduction for multi-label classification," in *Proc. IJCAI*, Jul. 2009, pp. 1077–1082.

[22] Q. Gu, Z. Li, and J. Han, "Joint feature selection and subspace learning," in *Proc. Int. Joint Conf. Artif. Intell.*, Jul. 2011, pp. 1294–1299.

[23] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

[24] T. Rainforth and F. Wood. (2015). "Canonical correlation forests." [Online]. Available: https://arxiv.org/abs/1507.05444

[25] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[26] J. Xia, N. Yokoya, and A. Iwasaki, "Hyperspectral image classification with canonical correlation forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 421–431, Jan. 2017.
[27] N. Yokoya *et al.*, "Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1363–1377, May 2018.
[28] D. Hong, N. Yokoya, J. Xu, and X. Zhu, "Joint & Progressive learning from high-dimensional data for multi-label classification," in *Proc. ECCV*, 2018, pp. 469–484.
[29] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
[30] X. He and P. Niyogi, "Locality preserving projections," in *Proc. NIPS*, Dec. 2004, pp. 153–160.
[31] C. Wang and S. Mahadevan, "A general framework for manifold alignment.," in *Proc. AAAI*. Nov. 2009, pp. 53–58.
[32] 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.
[33] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over Chikusei," Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27, 2016.
[34] R. Lai and S. Osher, "A splitting method for orthogonality constrained problems," *J. Sci. Comput.*, vol. 58, no. 2, pp. 431–449, 2014.

Danfeng Hong (S'16) received the B.Sc. degree in computer science and technology from the Neusoft College of Information, Northeastern University, Shenyang, China, in 2012, and the M.Sc. degree in computer vision from Qingdao University, Qingdao, China, in 2015. He is currently pursuing the Ph.D. degree in signal processing in earth observation from the Technical University of Munich, Munich, Germany, and the Remote Sensing Technology Institute, German Aerospace Center (DLR), Weßling, Germany.

His research interests include signal/image processing and analysis, pattern recognition, machine/deep learning and their applications in Earth Vision.

Naoto Yokoya (S'10–M'13) received the M.Sc. and Ph.D. degrees in aerospace engineering from The University of Tokyo, Tokyo, Japan, in 2010 and 2013, respectively.

From 2012 to 2013, he was a Research Fellow with the Japan Society for the Promotion of Science, Tokyo. From 2013 to 2017, he was an Assistant Professor with The University of Tokyo. From 2015 to 2017, he was also an Alexander von Humboldt Research Fellow with the German Aerospace Center (DLR), Weßling, Germany, and Technical University of Munich, Munich, Germany. Since 2018, he has been leading the Geoinformatics Unit with the RIKEN Center for Advanced Intelligence Project, RIKEN, Tokyo. His research interests include image analysis and data fusion in remote sensing.

Dr. Yokoya is a Co-Chair of IEEE Geoscience and Remote Sensing Image Analysis and Data Fusion Technical Committee since 2017. In 2017, he was a recipient of the Data Fusion Contest 2017 organized by the Image Analysis and Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society. His model was the most accurate among over 800 submissions.

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

Dr. Chanussot was a member of the IEEE Geoscience and Remote Sensing Society AdCom from 2009 to 2010, in charge of membership development and Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008. He is a member of the Institut Universitaire de France from 2012 to 2017. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing. He was the Chair from 2009 to 2011 and the Co-Chair of the GRS Data Fusion Technical Committee from 2005 to 2008. He was the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He is the Founding President of IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010 which received the 2010 IEEE GRSS Chapter Excellence Award. He was a co-recipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 and 2015 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award, and the IEEE GRSS 2013 Highest Impact Paper Award. He was an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2005 to 2007 and *Pattern Recognition* from 2006 to 2008. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. Since 2007, he has been an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and since 2018, he has also been an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING. He was the Guest Editor for the PROCEEDINGS OF THE IEEE in 2013 and *IEEE Signal Processing Magazine* in 2014.

Xiao Xiang Zhu (S'10–M'12–SM'14) received the M.Sc., Dr.-Ing., and the Habilitation degrees in signal processing from the Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

She is currently the Professor of Signal Processing in Earth Observation, TUM, and German Aerospace Center (DLR), Weßling, Germany; the Head of the Department EO Data Science with the DLR's Earth Observation Center; and the Head of the Helmholtz Young Investigator Group SiPEO with DLR and TUM. She was a Guest Scientist or Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy; Fudan University, Shanghai, China; The University of Tokyo, Tokyo, Japan; and the University of California at Los Angeles, Los Angeles, CA, USA, in 2009, 2014, 2015, and 2016, respectively. Her main research interests include remote sensing and earth observation, signal processing, machine learning, and data science, with a special application focus on global urban mapping.

Dr. Zhu is a member of young academy (Junge Akademie/Junges Kolleg) at the Berlin-Brandenburg Academy of Sciences and Humanities and the German National Academy of Sciences Leopoldina and the Bavarian Academy of Sciences and Humanities. She is an Associate Editor of theIEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.