

Graph Regularized Nonlinear Ridge Regression for Remote Sensing Data Analysis

Renlong Hang, Qingshan Liu, *Senior Member, IEEE*, Huihui Song, Yubao Sun, Fuping Zhu, and Hucheng Pei

Abstract—In this paper, a graph regularized nonlinear ridge regression (RR) model is proposed for remote sensing data analysis, including hyper-spectral image classification and atmospheric aerosol retrieval. The RR is an efficient linear regression method, especially in handling cases with a small number of training samples or with correlated features. However, large amounts of unlabeled samples exist in remote sensing data analysis. To sufficiently explore the information in unlabeled samples, we propose a graph regularized RR (GRR) method, where the vertices denote labeled or unlabeled samples and the edges represent the similarities among different samples. A natural assumption is that the predict values of neighboring samples are close to each other. To further address the nonlinearly separable problem in remote sensing data caused by the complex acquisition process as well as the impacts of atmospheric and geometric distortions, we extend the proposed GRR into a kernelized nonlinear regression method, namely KGRR. To evaluate the proposed method, we apply it to both classification and regression tasks and compare with representative methods. The experimental results show that KGRR can achieve favorable performance in terms of predictability and computation time.

Index Terms—Atmospheric aerosol retrieval, feature selection, graph regularization, hyperspectral image classification, kernel extension, ridge regression (RR).

I. INTRODUCTION

CURRENT remote sensors can fully portray the earth surface through tens of or even hundreds of contiguous and narrow spectral bands. The resulting multispectral or hyperspectral images capture rich information of land covers. Such high-dimensional datasets pose many challenging problems. First, with a limit number of training samples, the high-dimensional remote sensing datasets easily lead to the well-known Huge phenomenon. This means that, when the number of features exceeds a threshold, the classification accuracy starts to decrease. Second, high correlations exist in these high-dimensional datasets

Manuscript received December 1, 2015; revised March 25, 2016; accepted May 17, 2016. Date of publication June 21, 2016; date of current version December 22, 2016. This work was supported by the Natural Science Foundation of China under Grants 41501377, 61300162, and 61532009, and in part by the Foundation of Jiangsu Province, China, under Grants BK20150906, BK20131003, 15KJB170012, and 15KJA520001. (*Corresponding author: Qingshan Liu.*)

R. Hang, Q. Liu, H. Song, and Y. Sun are with the Jiangsu Key Laboratory of Big Data Analysis Technology, Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: renlong_hang@163.com; qslu@nuist.edu.cn; songhuihui@nuist.edu.cn; sunyb@nuist.edu.cn).

F. Zhu and H. Pei are with the Beijing Electro-Mechanical Engineering Institute, Beijing 100083, China (e-mail: zhufuping@126.com; Peihch@163.com).

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.2016.2574802

due to the dense sampling in spectral domain. These redundant information may not improve the performance of models, but increase computation and storage cost. Finally, complex atmospheric transmission and interference makes some bands contain less information or even be corrupted by noise.

Many approaches have been proposed to address the aforementioned issues. One intuitive and effective preprocessing method is dimensionality reduction [1]–[3]. However, most of the existing dimensionality reduction methods ignore the interaction with predictive models, leading to the suboptimal features for the models. Another widely adopted method is to employ a support vector machine (SVM) model [4]–[6] due to its lower sensitivity to the curse of dimensionality and sparse representation of the decision function as compared to traditional models. Besides dimensionality reduction and SVM methods, another promising scheme is regularization methods [7]–[9], which incorporate feature selection or transformation as part of the model training process. Generally, the objective function of regularization methods consists of an approximate term and a regularization term. With regard to the second term, researchers proposed various methods, among which \mathcal{L}_1 -norm [10]–[12] and \mathcal{L}_2 -norm [13]–[15] are two of the most popular ones. For these two regularizations, the Lasso and ridge regression (RR) are two representative models, respectively. The discussions in [7] and [9] demonstrate that the prediction performance of the Lasso is dominated by RR when there are high correlations among features. From this point of view, RR is more suitable for remote sensing data analysis than the Lasso.

In this paper, we focus on the regularization-based method, particularly RR. It is a well-known multiple linear regression technique by incorporating \mathcal{L}_2 -norm regularization into the ordinary least square [16]. Due to its flexibility and simplicity, it has been widely applied in various domains such as face recognition [17], bioinformatics [18] and cheminformatics [19]. Recently, some researchers attempted to employ the RR model in biophysical parameter retrieval from remote sensing data [20], [21]. In the field of remote sensing, there often exist large amounts of unlabeled samples. To fully explore the wealth of unlabeled samples, we propose a graph-based semisupervised learning (SSL) model, namely graph regularized RR (GRR). Specifically, we first construct an undirected and edge-weighted graph, where the vertices denote labeled or unlabeled samples and the edges represent the similarities among different samples. Then, the graph information is incorporated into the original objective function of RR as a regularizer. Finally, considering that most of remote sensing datasets are nonlinearly distributed, a generalized kernel version of GRR (KGRR) is further derived. It is worth noting that most of the existing graph-based SSL

methods are evaluated on remote sensing image classification [22]–[24], but the proposed method will be evaluated on both classification and regression.

The rest of this paper is outlined as follows. Section II introduces the proposed method in detail. Section III presents the datasets used and the experimental results in both classification and regression problems, followed by the conclusion in Section IV.

II. METHODOLOGY

Suppose given l labeled samples $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^l$ and u unlabeled samples $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$, with $\mathbf{x}_i \in R^d$ where d is the dimension of features. For classification problems, we adopt a 1-of- C encoding method, where C is the number of classes, to convert the original label into a C -dimensional vector representation as in [25] and [26]. In particular, if \mathbf{x}_i belongs to the j th class, $y_{ij} = 1$; otherwise, $y_{ij} = 0$. For regression problems, we just use the original labels, because it is impossible to encode continuous dependent variables. In matrix format, the training set can be represented as $\mathbf{X}_t = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l]^\top$, and the whole dataset as $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_l, \dots, \mathbf{x}_{l+u}]^\top$, where the superscript \top denotes the transpose of matrix.

A. GRR

The main idea of linear regression is to fit a function $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{w} + b$ such that the residual sum of square (loss function) is minimized: $(\mathbf{w}^*, b^*) = \arg \min_{\mathbf{w}, b} \sum_{i=1}^l \|f(\mathbf{x}_i) - \mathbf{y}_i\|^2$ [27]. For simplicity, we append a new element 1 to each \mathbf{x}_i . Then, the coefficient b can be absorbed into \mathbf{w} . Let $\mathbf{Y}_t = [\mathbf{y}_1, \dots, \mathbf{y}_l]^\top$, we can rewrite the loss function as a matrix form $\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{X}_t \mathbf{w} - \mathbf{Y}_t\|_F^2$ whose solution is $\mathbf{w}^* = (\mathbf{X}_t^\top \mathbf{X}_t)^{-1} \mathbf{X}_t^\top \mathbf{Y}_t$. However, in many practical applications, the covariance matrix $\mathbf{X}_t^\top \mathbf{X}_t$ is singular, because the features are highly correlated with each other, or the number of features is greater than that of samples. A popular solution is to impose a penalty on the norm of \mathbf{w} .

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{X}_t \mathbf{w} - \mathbf{Y}_t\|_F^2 + \alpha \|\mathbf{w}\|^2 \quad (1)$$

where α is a regularization parameter. The solution of (1) is $\mathbf{w}^* = (\mathbf{X}_t^\top \mathbf{X}_t + \alpha \mathbf{I})^{-1} \mathbf{X}_t^\top \mathbf{Y}_t$, where \mathbf{I} is a $d \times d$ identity matrix. The term $\|\mathbf{w}\|^2$ is called the Tikhonov regularizer [8]. In statistics, this regularization method is called RR.

In the field of remote sensing, it is difficult or expensive to acquire the ground-truth information. Taking atmospheric aerosol retrieval as an example, the retrievals of ground-based instruments are often used as a proxy of the true aerosol values [28]–[30]. There are only dozens of ground-based stations in China, resulting in a limit number of ground truth. Besides, the ground-based observations must match with the satellite observations in both time and space, which further reduces the available ground truth. To address this issue, many approaches have been proposed to fully exploit the rich information of unlabeled samples. Graph model is an effective method by linking labeled and unlabeled samples via vertices and edges. We thus propose to introduce a graph-based regularization term into RR,

named as GRR. Suppose we build a graph G with $l + u$ vertices, where each vertex corresponds to a sample, and the similarity between vertices i and j is denoted as S_{ij} , then a symmetric matrix \mathbf{S} is derived whose element S_{ij} is defined as

$$S_{ij} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}}, & \text{if } \mathbf{x}_i \in N_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_k(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where $N_k(\mathbf{x}_j)$ is the set of k -nearest neighbors of \mathbf{x}_j and σ is the parameter of the Heat kernel. A natural assumption of the graph regularizer is that nearby points in the graph G have similar labels, which motivates the choice of the following quadratic energy function:

$$f^* = \arg \min_f \sum_{i,j=1}^{l+u} \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|^2 S_{ij} = \arg \min_f \mathbf{f}^\top \mathbf{L} \mathbf{f} \quad (3)$$

where $\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_l), \dots, f(\mathbf{x}_{l+u})]$ and \mathbf{L} is the graph Laplacian matrix. Using some simple algebraic formulations, we have $\mathbf{L} = \mathbf{D} - \mathbf{S}$, where \mathbf{D} is a diagonal matrix whose entries equal to the column or row sum of \mathbf{S} : $D_{ii} = \sum_j S_{ji}$. Incorporating (3) into (1), the final objective function of GRR is

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\mathbf{X}_t \mathbf{w} - \mathbf{Y}_t\|_F^2 + \alpha \|\mathbf{w}\|^2 + \beta \mathbf{w}^\top \mathbf{X}^\top \mathbf{L} \mathbf{X} \mathbf{w}. \quad (4)$$

Obviously, (4) is a convex function and has a global optimal solution $\mathbf{w}^* = (\mathbf{X}_t^\top \mathbf{X}_t + \alpha \mathbf{I} + \beta \mathbf{X}^\top \mathbf{L} \mathbf{X})^{-1} \mathbf{X}_t^\top \mathbf{Y}_t$. It is worth noting that RR is a special case of GRR when β equals to 0.

B. Kernelized GRR

GRR assumes that there exists a linear relation between \mathbf{X} and \mathbf{Y} . However, nonlinearities are often exhibited in remote sensing data due to the effects of multipath scattering, variations in sun-canopy-sensor geometry, nonhomogeneous composition of pixels, and attenuating properties of media [31]. To properly fit the remote sensing data into the GRR scheme, we employ kernel methods, which can effectively transform the data from a nonlinear space into a linear space [32], [33]. Taking Fig. 1 as an example, the original data points from two classes of the selected image patch are linearly nonseparable in spectral feature space due to overlapping areas. But after mapping them to another higher dimensional space via kernel tricks, they can be linearly separated. Thus, we further extend the proposed GRR into a kernelized GRR as follows.

Kernel methods aim to project the data into a Hilbert space \mathcal{H} with very high or even infinite dimension, and then, perform a linear algorithm in \mathcal{H} . Given a nonlinear mapping function $\phi(\mathbf{x}) : R^d \rightarrow \mathcal{H}$, the projected training set and the whole dataset can be represented as $\phi_t = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_l)]^\top$ and $\phi = [\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_l), \dots, \phi(\mathbf{x}_{l+u})]^\top$, respectively. Besides, according to *representer theorem* [32], $\mathbf{w} = \sum_{i=1}^l \phi(\mathbf{x}_i) a_i = \phi_t^\top \mathbf{A}$, where $\mathbf{A} = [a_1, \dots, a_i, \dots, a_l]^\top$. Substituting them into

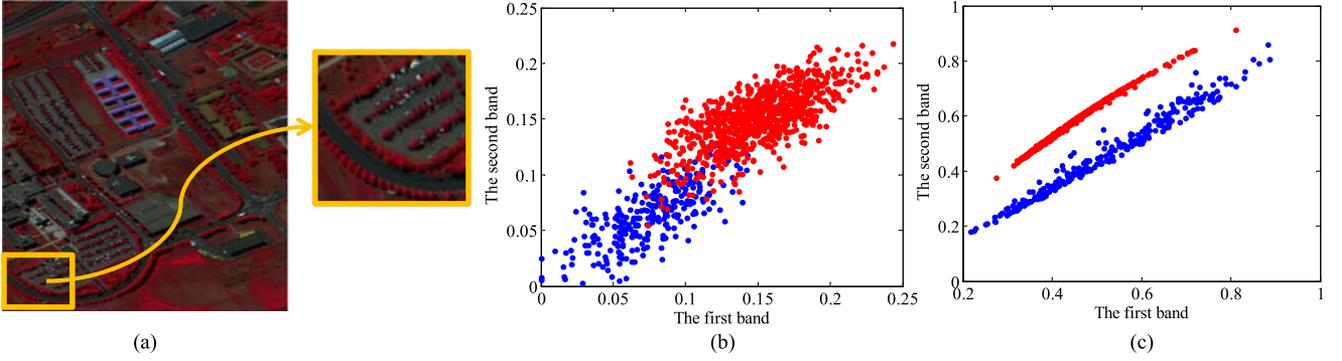


Fig. 1. Illustration of kernel operator. (a) Color hyperspectral image over Pavia, northern Italy. (b) Original two-classes data distribution of the selected patch. (c) Higher dimensional projection of the original data via Gaussian kernel.

(4), the objective function of KGRR is formulated as

$$\begin{aligned} \mathbf{A}^* = \arg \min_{\mathbf{A}} & \|\phi_t \phi_t^\top \mathbf{A} - \mathbf{Y}_t\|_F^2 + \alpha \mathbf{A}^\top \phi_t \phi_t^\top \mathbf{A} \\ & + \beta \mathbf{A}^\top \phi_t \phi_t^\top \mathbf{L} \phi_t \phi_t^\top \mathbf{A}. \end{aligned} \quad (5)$$

Nonetheless, directly computing ϕ is nontrivial, but we can get around it by calculating the dot product in high-dimensional space via kernel trick, which can be expressed as $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, where the operator $\langle \cdot \rangle$ means inner product, and $K(\cdot)$ denotes kernel function. Therefore, we rewrite (5) as

$$\mathbf{A}^* = \arg \min_{\mathbf{A}} \|\mathbf{K}_t \mathbf{A} - \mathbf{Y}_t\|_F^2 + \alpha \mathbf{A}^\top \mathbf{K}_t \mathbf{A} + \beta \mathbf{A}^\top \mathbf{K} \mathbf{L} \mathbf{K}^\top \mathbf{A}. \quad (6)$$

Let the derivative of $L(\mathbf{A})$ with respect to \mathbf{A} equal to $\mathbf{0}$, we get $\mathbf{A}^* = (\mathbf{K}_t^\top \mathbf{K}_t + \alpha \mathbf{K}_t + \beta \mathbf{K} \mathbf{L} \mathbf{K}^\top)^{-1} \mathbf{K}_t^\top \mathbf{Y}_t$. For any unlabeled data point \mathbf{x}_j , its predicted value is $f(\mathbf{x}_j) = \sum_{i=1}^l \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle a_i = \sum_{i=1}^l K(\mathbf{x}_i, \mathbf{x}_j) a_i$.

III. EXPERIMENTS

We evaluate the proposed KGRR algorithm in two remote sensing tasks: atmospheric aerosol retrieval and hyperspectral image classification. As a regression problem in the first task, the original values in the label vector remain unchanged. For the classification problem in the latter task, we adopt a 1-of- C encoding method to convert the original label into a C -dimensional vector representation.

A. Atmospheric Aerosol Retrieval

1) *Datasets*: Atmospheric aerosol retrieval can be addressed as a regression problem to learn a mapping from multi-band spectral values to aerosol parameters (e.g., aerosol optical depth (AOD) in this paper). As in [28]–[30], we use Level 2.0 AERONET retrievals as the ground truth (targets). AERONET is a global aerosol observing network with about 250 ground-based instruments [34]. Most of these stations measure AOD in different spectral bands centered around the nominal wavelengths of 340, 380, 440, 670 nm, and others [35]. To facilitate intercomparisons with measurements from remote sensing data, the AERONET retrievals are first interpolated to 550 nm using the quadratic fit on log–log scale for all wavelengths [36]. For

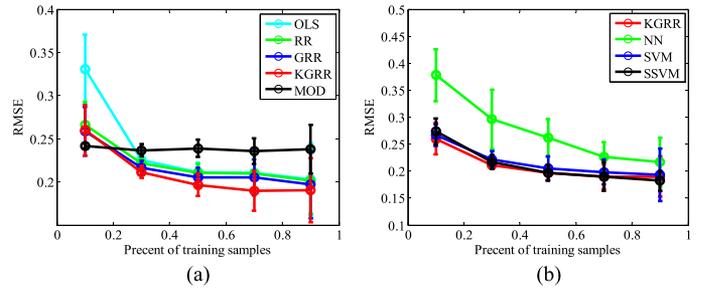


Fig. 2. RMSE and standard deviations achieved by eight different methods under different number of training samples on MODIS dataset. (a) Comparisons among KGRR and four other methods. (b) Comparisons among KGRR, NN, SVM and SSVM.

the inputs of the proposed algorithm, we use the spectral values along all bands from the following two remote sensors.

The first is moderate resolution imaging spectroradiometer (MODIS), which is a key instrument aboard the TERRA satellite. With a single camera, MODIS observes the top-of-the-atmosphere reflectance over 36 spectral bands between 410 nm and $14 \mu\text{m}$ [37]. We obtain the MODIS Level-1B calibrated radiance product MOD021KM with spatial resolution of 1 km from the Beijing AERONET location between January 2002 and December 2014. Thereafter, Level 2.0 AERONET data are collocated in space and time with MODIS data. The detailed process can be found in [28]. We obtain a total of 843 spatially and temporally collocated observations from MODIS and AERONET.

The second sensor is multiangle imaging spectroradiometer (MISR). The MISR is one of the five instruments mounted on TERRA spacecraft. MISR consists of nine pushbroom cameras arranged in different view angles relative to the earth’s surface. Each camera covers 360-km wide swath and provides four spectral bands in blue, green, red, and near infrared (NIR) that are centered at 443, 555, 670, and 865 nm, respectively. The resolution of all bands in nadir view and the red bands at all nine angles is 275 m and the resolution of other bands is 1.1 km. We download 1045 collocated MISR and AERONET data from multisensor aerosol products sampling system [35], covering the whole 23 stations at all available time in China.

2) *Experimental Setup*: To demonstrate the superiority of the proposed KGRR model on aerosol retrieval. We compare it with the following seven methods: 1) the ordinary least-square regression (OLS), which is a special case of GRR when setting $\alpha = \beta = 0$; 2) RR, which is also a special case of GRR when setting $\beta = 0$; 3) GRR; 4) the operational satellite retrieval algorithms based on physical models, namely MOD and MISR, respectively; 5) an SVM [30]; 6) neural network (NN) [28]; and 7) the semisupervised SVM with graph-based kernel (SSVM) proposed in [38]. For NN, the optimal number of hidden nodes is chosen from [2, 50] in steps of 5 via a fivefold cross validation. For an SVM, we adopt the Gaussian kernel since it usually achieves the best results compared to other kernels. The optimal variance parameter σ for Gaussian kernel and the regularization parameter C in SVM are both selected from $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ via a fivefold cross validation. For the two regularization parameters of GRR and KGRR, we also choose them via a fivefold cross validation from the given set $\{10^{-3}, 10^{-2}, \dots, 10^3\}$. Besides, the physical models attempt to take into account numerous physical variables affecting the radiometric characteristics of remote sensing data, such as atmospheric conditions, solar azimuth and zenith angles, sensor azimuth, and zenith angles, etc. Complex mathematical formulations are set up to represent the relationships between these variables according to radiation transfer equation. To simplify the radiative transfer calculations, a lookup table (LUT) is used to simulate the radiative properties of the atmosphere calculated for expected aerosol types at particular wavelengths, angles and aerosol loading. Spectral reflectance from the LUT is then compared with the satellite-observation value to find the best match, and the corresponding AOT is the final retrieval result. In all experiments, we randomly divide the whole dataset into two groups: training set and testing set. The training set is used to train all of the machine learning models, while the testing set is used to evaluate the performance of each model. In order to reduce the effects of random selection, all the algorithms are repeated ten times and the average performance is reported. Without loss of generality, we use two mainstream evaluation metrics: the root-mean-square error (RMSE) and Pearson's correlation coefficient R , which are used to evaluate the accuracy of the estimations and the goodness of fit, respectively.

3) *Results and Discussion*: For MODIS data, the RMSE and standard deviations achieved by five different methods under different numbers of training samples are shown in Fig. 2(a), from which several conclusions can be drawn. First, RR is better than OLS when 10% training samples are available, owing to singular covariance matrix in OLS. However, when training samples are more than 30%, the covariance matrix is nonsingular and the performance of OLS is close to RR. Second, KGRR and GRR are better than RR and OLS all the time, because they are able to take advantage of unlabeled samples, which certifies the efficiency of graph regularization. Third, due to the nonlinear distribution of data, KGRR yields the best results especially when training samples are more than 30%. The last but not the least, with the increasing number of training samples, RMSE values decrease for machine learning models, while that is stable for the physical model. In particular, when there

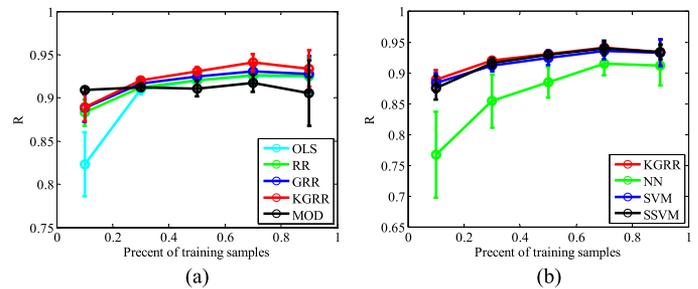


Fig. 3. R and standard deviations achieved by eight different methods under different number of training samples on MODIS dataset. (a) Comparisons among KGRR and four other methods. (b) Comparisons among KGRR, NN, SVM and SSVM.

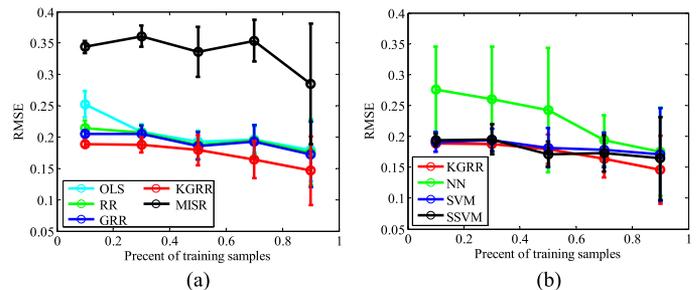


Fig. 4. RMSE and standard deviations achieved by eight different methods under different number of training samples on MISR dataset. (a) Comparisons among KGRR and four other methods. (b) Comparisons among KGRR, NN, SVM and SSVM.

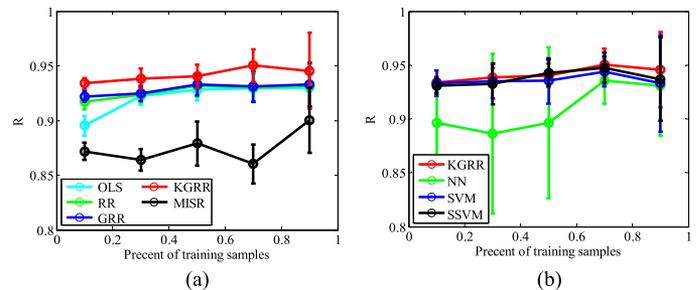


Fig. 5. R and standard deviations achieved by eight different methods under different number of training samples on MISR dataset. (a) Comparisons among KGRR and four other methods. (b) Comparisons among KGRR, NN, SVM and SSVM.

are small amounts of training samples (i.e., 10%), the physical model gets a higher performance than machine learning models. On the contrary, given enough training samples (more than 30%), machine learning models significantly outperform the physical model. This indicates that the performance of machine learning methods depends on the number of training samples. Similar conclusions can be drawn from another indicator in Fig. 3(a). Besides, Figs. 2(b) and 3(b) demonstrate the comparison results among KGRR and three popular machine learning retrieval models: SVM, NN, and SSVM. The proposed KGRR remarkably outperforms NN and is a little better than SVM and SSVM. Nevertheless, KGRR is much faster than SVM under the same number of training samples as shown in Fig. 6(a).

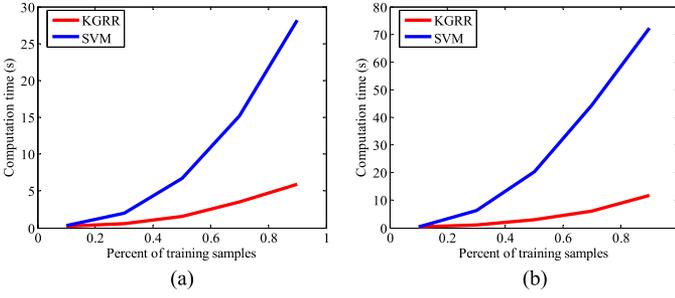


Fig. 6. Computation time comparisons between KGRR and SVM on two different datasets. (a) MODIS dataset. (b) MISR dataset.

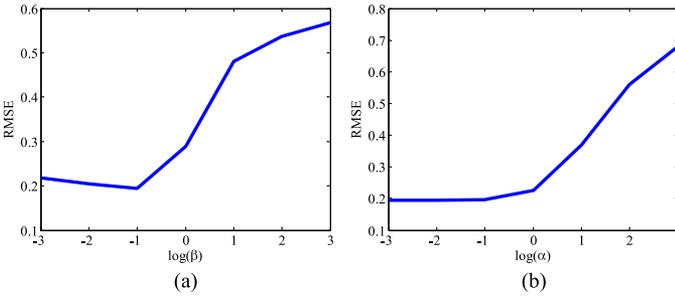


Fig. 7. RMSE results of KGRR versus two regularization parameters on MISR dataset. (a) β . (b) α .

For MISR data, Figs. 4 and 5 show RMSE and R values obtained by eight different models, respectively. Similar conclusions can be drawn as on MODIS data, i.e., machine learning models perform much better than the physical model and KGRR achieves the best results. Besides, Figs. 4(b), 5(b), and 6(b) demonstrate that KGRR can do as well as, if not better than an SVM with much less time costs when training samples are less than 50%.

4) *Parameter Analysis*: Two important parameters need to be tuned in KGRR: the regularization parameters α and β . To examine their effects on regression results, we conduct experiments on the MISR dataset. For simplicity, we fix α when tune β and vice versa. Fig. 7(a) demonstrates the RMSE results against β . Obviously, the RMSE results decrease as β increases when β is smaller than 10^{-1} . In contrast, when β is larger than 10^{-1} , the RMSE results increase as β increases. Therefore, the optimal β equals to 10^{-1} . Similarly, the RMSE results versus α is shown in Fig. 7(b), which indicates that RMSE values are stable when α is small while RMSE increases when α increases. Thus, the best α is set to 10^{-3} .

B. Hyperspectral Image Classification

1) *Datasets*: To examine the efficiency of KGRR on remote sensing image classification, we apply it to two hyperspectral datasets acquired by different sensors. The first is Indian Pines (IP) dataset, which was acquired by the AVIRIS sensor over the Indian Pine test site in northwestern Indiana, USA, on June 12, 1992. The original dataset contains 224 spectral bands. But we utilize 200 of them after removing four bands containing zero values and 20 noisy bands affected by water absorption. The

spatial size of the image is 145×145 pixels, and the spatial resolution is 20 m. The false-color composite image and the ground-truth map are shown in Fig. 8(a) and (b). The available number of samples for each class shown in Table I ranges largely from 20 to 2455, which makes the classification task very difficult.

The second dataset is Pavia University Scene (PUS), which was acquired by the ROSIS sensor during a flight campaign over Pavia, northern Italy, on July 8, 2002. The original image is recorded with 115 spectral channels ranging from 0.43 to $0.86 \mu\text{m}$ and covering the visible and infrared spectrum. The utilized image contains 103 bands after removing noisy bands. The image size is 610×340 pixels with a spatial resolution of 1.3 m. A false-color composite image and the ground-truth map are shown in Fig. 8(c) and (d). In the ground-truth map, there are nine classes of land covers with more than 1000 labeled pixels for each class, which is listed in Table II.

2) *Experimental Setup*: To assess the performance of KGRR. We compare it with OLS, RR, GRR, SVM, and semisupervised SVM using the combination of Gaussian kernel and cluster kernel (SVM_Cluster) proposed in [39]. For regression models OLS, RR, GRR, and KGRR, we adopt the “winner-takes-all” strategy for classification problem as in [25]. The optimal regularization parameters α and β for GRR and KGRR are both chosen from $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ via a fivefold cross validation. For an SVM model, we exploit two different kernels: linear kernel and Gaussian kernel. For sake of simplicity, we represent them as SVM_Gaussian and SVM_linear, respectively. The best variance parameter σ for Gaussian kernel and regularization parameter C in SVM are both selected from $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ via a fivefold cross validation. In all experiments, we randomly select 1% samples from each class as the training set and the rest as the testing set. The detailed numbers are demonstrated in Table I and Table II, respectively. In order to reduce the effects of random selection, all algorithms are repeated ten times and the average results are reported. The classification performance is evaluated by the overall accuracy (OA), the average accuracy (AA), the per-class accuracy, and the Kappa coefficient κ . OA is defined by the ratio between the number of correctly classified pixels to the total number of pixels in the testing set. AA refers to the average of accuracies in all classes, and κ is the percentage of agreement corrected by the number of agreements that would be expected purely by chance.

3) *Results and Discussion*: For IP dataset, the classification results achieved by different models are demonstrated in Table III, where bold fonts indicate the best results. From this table, we observe that OLS does not work in the case with a small number of training samples and high-dimensional input features. This problem can be alleviated by adding a Tikhonov regularizer to OLS, i.e., RR model. GRR is better than OLS and RR, because it is able to exploit the wealth of unlabeled samples. Nevertheless, it is inferior to SVM models, including SVM_Gaussian and SVM_linear, in terms of OA and κ , and superior to them in terms of AA. This can be explained by that Gaussian kernel makes the data linearly separated in very

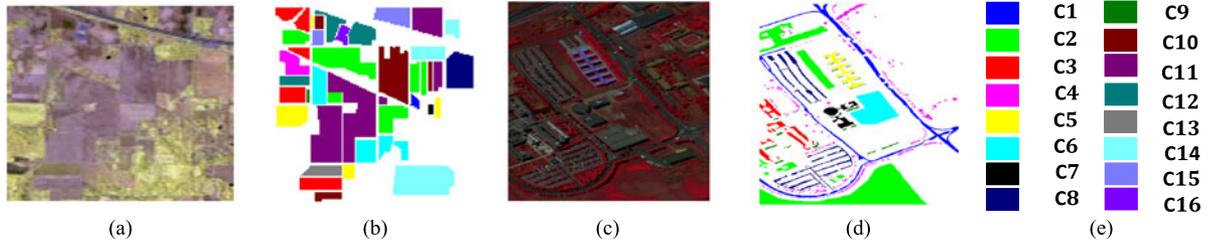


Fig. 8. RGB composite images and ground-truth maps on two datasets. (a) and (b) IP dataset. (c) and (d) PUS dataset. (e) Class labels for different classes.

TABLE I
NUMBER OF PIXELS FOR TRAINING/TESTING AND THE TOTAL NUMBER OF PIXELS FOR EACH CLASS IN THE IP GROUND-TRUTH MAP

Label	Class Name	Total	Sample Training	Sample Testing	Label	Class Name	Total	Sample Training	Sample Testing
C1	Alfalfa	46	1	45	C9	Oats	20	1	19
C2	Corn-notill	1428	15	1413	C10	Soybean-notill	972	10	962
C3	Corn-mintill	830	9	821	C11	Soybean-mintill	2455	25	2430
C4	Corn	237	3	234	C12	Soybean-clean	593	6	587
C5	Grass-pasture	483	5	478	C13	Wheat	205	3	202
C6	Grass-trees	730	8	722	C14	Woods	1265	13	1252
C7	Grass-pasture-mowed	28	1	27	C15	Buildings-Grass-Trees-Drives	386	4	382
C8	Hay-windrowed	478	5	473	C16	Stone-Steel-Towers	93	1	92

TABLE II
NUMBER OF PIXELS FOR TRAINING/TESTING AND THE TOTAL NUMBER OF PIXELS FOR EACH CLASS IN THE PUS GROUND-TRUTH MAP

Label	Class Name	Total	Sample Training	Sample Testing	Label	Class Name	Total	Sample Training	Sample Testing
C1	Asphalt	6631	67	6564	C6	Bare Soil	5029	51	4978
C2	Meadows	18649	187	18462	C7	Bitumen	1330	14	1316
C3	Gravel	2099	21	2078	C8	Self-Blocking Bricks	3682	37	3645
C4	Trees	3064	31	3033	C9	Shadows	947	10	937
C5	Painted metal sheets	1345	14	1331					

TABLE III
OA, AA, PER-CLASS ACCURACY (%), κ , AND STANDARD DEVIATION OF TEN RUNS ACHIEVED BY SEVEN METHODS ON IP DATASET

Label	OLS	RR	GRR	SVM_Gaussian	SVM_Linear	SVM_Cluster	KGRR
C1	1.85 ± 4.01	13.59 ± 31.27	22.54 ± 41.16	17.05 ± 13.74	20.80 ± 19.29	25.79 ± 19.73	24.16 ± 16.87
C2	17.27 ± 6.09	48.51 ± 3.12	50.08 ± 3.67	49.11 ± 4.62	44.59 ± 5.34	42.81 ± 7.09	52.55 ± 4.76
C3	11.00 ± 5.67	32.94 ± 8.21	34.95 ± 9.49	47.99 ± 25.31	38.65 ± 15.49	49.96 ± 17.38	43.89 ± 8.93
C4	2.94 ± 1.94	25.13 ± 17.54	27.21 ± 16.93	27.44 ± 15.15	26.35 ± 12.71	32.66 ± 14.63	34.90 ± 8.53
C5	14.41 ± 12.29	67.67 ± 18.53	68.53 ± 19.01	43.87 ± 23.87	57.65 ± 23.88	72.85 ± 14.73	64.91 ± 16.93
C6	24.43 ± 21.26	72.62 ± 5.27	71.75 ± 4.87	72.36 ± 8.69	71.86 ± 5.71	73.27 ± 4.43	71.78 ± 5.22
C7	0.34 ± 0.68	44.87 ± 37.74	48.33 ± 44.76	28.19 ± 20.89	31.66 ± 26.29	38.75 ± 34.08	43.82 ± 22.86
C8	23.53 ± 30.16	83.65 ± 3.65	83.38 ± 3.75	85.44 ± 3.84	84.38 ± 3.80	85.71 ± 5.89	85.30 ± 2.56
C9	0.32 ± 1.00	12.67 ± 17.13	17.29 ± 24.28	15.39 ± 21.64	16.55 ± 16.38	14.13 ± 9.01	27.80 ± 20.16
C10	11.19 ± 5.03	39.53 ± 5.16	41.79 ± 5.81	40.76 ± 18.21	41.38 ± 15.79	48.28 ± 5.77	46.76 ± 6.58
C11	22.66 ± 9.79	49.63 ± 2.17	49.68 ± 2.29	54.90 ± 7.13	58.66 ± 8.76	48.76 ± 9.14	58.68 ± 4.21
C12	10.29 ± 6.32	37.26 ± 3.37	39.10 ± 4.64	29.03 ± 15.98	39.34 ± 13.07	31.25 ± 11.27	37.36 ± 9.48
C13	17.27 ± 23.70	78.99 ± 9.38	80.19 ± 9.80	66.78 ± 24.71	75.07 ± 9.86	85.39 ± 9.11	75.76 ± 8.97
C14	26.59 ± 18.17	76.70 ± 3.79	77.55 ± 5.55	80.24 ± 5.59	81.52 ± 5.64	80.18 ± 6.47	81.01 ± 4.24
C15	5.95 ± 5.41	52.47 ± 15.98	51.33 ± 12.77	51.42 ± 18.06	41.50 ± 10.85	41.42 ± 16.13	51.23 ± 14.10
C16	0.43 ± 1.36	72.88 ± 38.87	73.97 ± 39.08	90.78 ± 26.89	90.84 ± 26.98	79.18 ± 41.79	90.92 ± 24.93
OA	13.43 ± 4.04	56.16 ± 1.22	57.16 ± 1.27	58.15 ± 2.18	57.33 ± 3.05	57.30 ± 3.91	59.91 ± 1.59
AA	11.91 ± 4.94	50.57 ± 2.47	52.35 ± 3.16	51.31 ± 5.88	50.04 ± 9.06	53.15 ± 7.43	55.67 ± 2.94
κ	2.82 ± 2.82	48.58 ± 1.48	49.30 ± 1.66	51.04 ± 3.38	50.73 ± 4.53	50.54 ± 5.39	53.80 ± 1.93

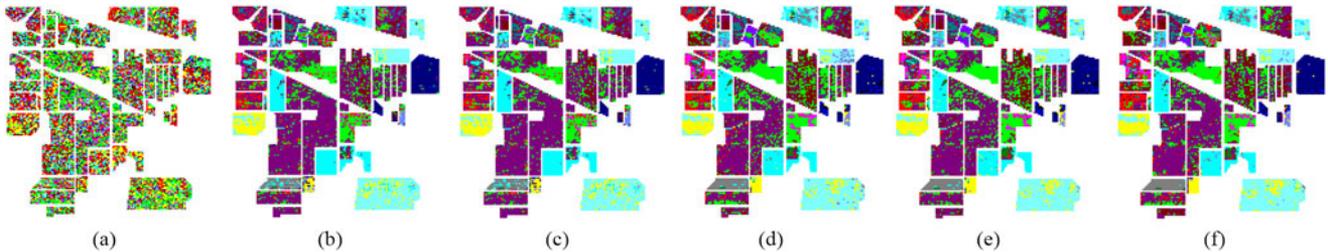


Fig. 9. Classification maps of six different methods on IP dataset. (a) OLS. (b) RR. (c) GRR. (d) SVM_Gaussian. (e) SVM_Linear. (f) KGRR.

TABLE IV
OA, AA, PER-CLASS ACCURACY (%), κ , AND STANDARD DEVIATION OF TEN RUNS ACHIEVED BY SEVEN METHODS ON PUS DATA SET

Label	OLS	RR	GRR	SVM_Gaussian	SVM_Linear	SVM_Cluster	KGRR
C1	59.17 ± 1.43	60.77 ± 1.29	61.13 ± 1.27	87.78 ± 2.76	84.95 ± 1.59	87.23 ± 3.36	88.05 ± 2.37
C2	80.92 ± 0.73	80.58 ± 0.46	79.58 ± 0.37	92.04 ± 1.41	89.99 ± 0.68	91.51 ± 0.67	91.21 ± 1.23
C3	56.95 ± 6.95	71.78 ± 6.79	78.35 ± 6.23	77.00 ± 4.50	75.18 ± 4.83	73.97 ± 3.06	75.91 ± 3.35
C4	91.62 ± 1.49	92.78 ± 1.61	92.47 ± 1.50	94.91 ± 2.04	91.97 ± 2.96	93.38 ± 3.50	96.44 ± 1.18
C5	100 ± 0	100 ± 0	100 ± 0	96.86 ± 1.68	96.33 ± 1.83	96.52 ± 2.69	98.53 ± 1.18
C6	53.50 ± 2.85	64.67 ± 3.05	69.19 ± 4.35	86.48 ± 4.93	83.78 ± 4.78	86.11 ± 2.37	90.38 ± 2.79
C7	37.02 ± 7.01	49.36 ± 18.62	38.59 ± 27.27	77.64 ± 6.59	69.54 ± 7.32	78.33 ± 5.14	76.04 ± 5.55
C8	48.01 ± 2.44	54.26 ± 2.85	55.55 ± 3.15	79.65 ± 2.46	80.53 ± 2.38	79.61 ± 1.12	78.05 ± 1.56
C9	73.27 ± 10.81	90.09 ± 7.42	98.32 ± 4.24	99.96 ± 0.06	99.97 ± 0.05	99.04 ± 2.15	99.98 ± 0.05
OA	72.02 ± 0.75	74.15 ± 0.64	75.06 ± 0.57	89.01 ± 0.63	87.02 ± 0.78	88.53 ± 0.59	89.17 ± 0.65
AA	66.72 ± 1.78	73.81 ± 2.14	74.80 ± 3.41	88.03 ± 1.40	85.81 ± 1.14	87.30 ± 0.69	88.29 ± 0.76
κ	61.20 ± 1.02	63.86 ± 0.90	64.55 ± 0.80	85.29 ± 0.93	82.57 ± 1.01	83.78 ± 0.79	85.31 ± 0.83

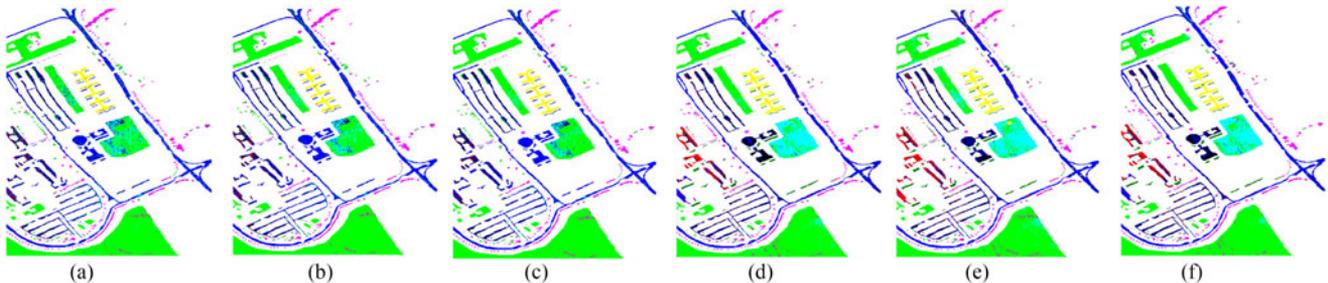
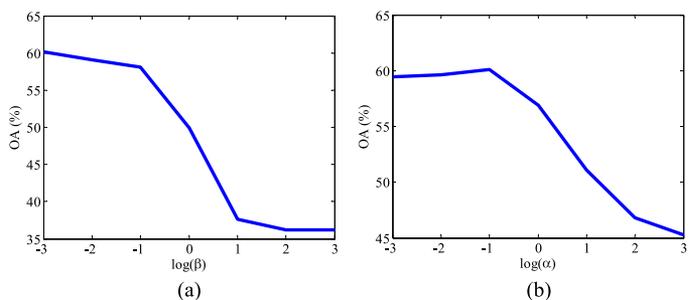


Fig. 10. Classification maps of six different methods on PUS dataset. (a) OLS. (b) RR. (c) GRR. (d) SVM_Gaussian. (e) SVM_Linear. (f) KGRR.

high-dimensional space. For a similar reason, as a kernel version of GRR, KGRR gets the best performance in terms of OA, AA, and κ . Besides, it is worth noting that the semisupervised method SVM_Cluster is inferior to SVM_Gaussian in terms of OA and κ , but superior to SVM_Gaussian in terms of AA. This means that some testing samples from the majority classes are misclassified, because they can be easily grouped into their neighboring clusters with a small number of samples during k -means clustering. Combining these improper clustering results with Gaussian kernel will inevitably degrade the overall performance. On the contrary, this problem does not exist in our proposed method, thereby leading to higher performance than SVM_Cluster. Fig. 9 shows the classification maps in one experiment among ten runs, where different colors correspond to different categories. These figures further verify the aforementioned conclusions.

Fig. 11. OA of KGRR versus two different regularization parameters on IP dataset. (a) β . (b) α .

For PUS dataset, Table IV reports the detailed classification results. Compared to IP dataset, the land covers in this data are easier to discriminate and the number of available samples is

significantly larger, leading to much better classification results than that on IP dataset. In particular, GRR achieves superior performance than OLS as well as RR, and inferior performance than SVM_Gaussian, SVM_linear, and SVM_Cluster in terms of OA, AA, and κ . More importantly, as a nonlinear extension of GRR, KGRR increases the OA from 87.02 to 89.17 compared with SVM_linear and is slightly better than SVM_Gaussian and SVM_Cluster. Fig. 10 demonstrates the classification maps, which further certifies the efficiency of KGRR.

4) *Parameter Analysis*: Similar to retrieval task, we also evaluate the effects of two regularization parameters α and β in KGRR on classification results using IP dataset. First, we fix α and check the effects of β . Fig. 11(a) shows that OA decreases as β increases. Therefore, the best β is 10^{-3} . Then, we fix β and alter α . Fig. 11(b) demonstrates that OA first increases, and then, decreases as α increases. Obviously, the optimal α equals to 10^{-1} .

IV. CONCLUSION

This paper proposed a graph regularized nonlinear RR model KGRR for different remote sensing applications. Compared to the widely used RR model, GRR can sufficiently exploit the wealth of unlabeled samples, thereby improving the predictability. Considering that most of remote sensing datasets are nonlinearly separable, a kernel extension of GRR, namely KGRR, was proposed. The basic idea is to map the original remote sensing data into a very high-dimensional feature space, in which the data can be linearly separated. To examine the effectiveness of KGRR, we applied it to atmospheric aerosol retrieval and hyperspectral image classification tasks on four datasets acquired by different sensors. The experimental results demonstrate that the proposed method KGRR performs better than OLS, RR, GRR, NN, SVM, SSVM, and SVM_Cluster. In addition, we also evaluated the influence of different regularization parameters α and β in KGRR on the final performance.

REFERENCES

- [1] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.
- [2] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.
- [3] S. Jia, Z. Zhu, L. Shen, and Q. Li, "A two-stage feature selection framework for hyperspectral image classification using few labeled samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1023–1035, Apr. 2014.
- [4] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [5] M. Pal and P. Mather, "Support vector machines for classification in remote sensing," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 1007–1011, 2005.
- [6] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.
- [7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., B, Methodol.*, vol. 58, pp. 267–288, 1996.
- [8] A. N. Tikhonov, "Regularization of incorrectly posed problems," *Sov. Math. Dokl.*, vol. 4, pp. 1624–1627, 1963.
- [9] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc., B, Statist. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.
- [10] P. Zhong, P. Zhang, and R. Wang, "Dynamic learning of SMLR for feature selection and classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 280–284, Apr. 2008.
- [11] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [12] H. Song, B. Huang, Q. Liu, and K. Zhang, "Improving the spatial resolution of landsat tm/etm+ through fusion with spot5 images via learning-based super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1195–1204, Mar. 2015.
- [13] X. Yu and S.-Y. Liong, "Forecasting of hydrologic time series with ridge regression in feature space," *J. Hydrol.*, vol. 332, no. 3, pp. 290–302, 2007.
- [14] H. Xue, Y. Zhu, and S. Chen, "Local ridge regression for face recognition," *Neurocomputing*, vol. 72, no. 4, pp. 1342–1346, 2009.
- [15] F. Douak, F. Melgani, and N. Benoudjit, "Kernel ridge regression with active learning for wind speed prediction," *Appl. Energy*, vol. 103, pp. 328–340, 2013.
- [16] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [17] S. An, W. Liu, and S. Venkatesh, "Face recognition using kernel ridge regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–7.
- [18] S. Giguère, M. Marchand, F. Laviolette, A. Drouin, and J. Corbeil, "Learning a peptide-protein binding affinity predictor with kernel ridge regression," *BMC Bioinf.*, vol. 14, no. 1, p. 82, 2013.
- [19] T. Hinkley *et al.*, "A systems analysis of mutational effects in hiv-1 protease and reverse transcriptase," *Nature Genetics*, vol. 43, no. 5, pp. 487–489, 2011.
- [20] J. Verrelst *et al.*, "Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for sentinel-2 and-3," *Remote Sens. Environ.*, vol. 118, pp. 127–139, 2012.
- [21] J. P. Rivera Caicedo, J. Verrelst, J. Muñoz-Marí, J. Moreno, and G. Camps-Valls, "Toward a semiautomatic machine learning retrieval of biophysical parameters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1249–1259, Apr. 2014.
- [22] G. Camps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
- [23] J. Bai, S. Xiang, and C. Pan, "A graph-based classification method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 803–817, Feb. 2013.
- [24] L. Yang, S. Yang, P. Jin, and R. Zhang, "Semi-supervised hyperspectral image classification using spatio-spectral Laplacian support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 651–655, Mar. 2014.
- [25] E. Izquierdo-Verdiguier, L. Gomez-Chova, L. Bruzzone, and G. Camps-Valls, "Semisupervised kernel feature extraction for remote sensing image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5567–5578, Sep. 2014.
- [26] J. Arenas-García and G. Camps-Valls, "Efficient kernel orthonormalized PLS for remote sensing applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2872–2881, Oct. 2008.
- [27] G. A. Seber and A. J. Lee, *Linear Regression Analysis*, vol. 936. New York, NY, USA: Wiley, 2012.
- [28] S. Vucetic, B. Han, W. Mi, Z. Li, and Z. Obradovic, "A data-mining approach for the validation of aerosol retrievals," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 113–117, Jan. 2008.
- [29] K. Ristovski, S. Vucetic, and Z. Obradovic, "Uncertainty analysis of neural-network-based aerosol retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 409–414, Feb. 2012.
- [30] D. Lary, L. Remer, D. MacNeill, B. Roscoe, and S. Paradise, "Machine learning and bias correction of MODIS aerosol optical depth," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 694–698, Oct. 2009.
- [31] C. M. Bachmann, T. L. Ainsworth, and R. A. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 441–454, Mar. 2005.
- [32] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [33] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
- [34] B. Holben *et al.*, "Aeronet/a federated instrument network and data archive for aerosol characterization," *Remote Sens. Environ.*, vol. 66, no. 1, pp. 1–16, 1998.

- [35] M. Petrenko, C. Ichoku, and G. Leptoukh, "Multi-sensor aerosol products sampling system (MAPSS)," *Atmospheric Meas. Tech.*, vol. 5, no. 5, pp. 913–926, 2012.
- [36] L. A. Remer *et al.*, "The MODIS aerosol algorithm, products, and validation," *J. Atmospheric Sci.*, vol. 62, no. 4, pp. 947–973, 2005.
- [37] V. V. Salomonson, W. Barnes, P. W. Maymon, H. E. Montgomery, and H. Ostrow, "Modis: Advanced facility instrument for studies of the earth as a system," *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 2, pp. 145–153, 1989.
- [38] G. Camps-Valls, J. Muñoz-Marí, L. Gómez-Chova, K. Richter, and J. Calpe-Maravilla, "Biophysical parameter estimation with a semisupervised support vector machine," *IEEE Geo. Remote Sens. Lett.*, vol. 6, no. 2, pp. 248–252, 2009.
- [39] D. Tuia and G. Camps-Valls, "Semisupervised remote sensing image classification with cluster kernels," *IEEE Geo. Remote Sens. Lett.*, vol. 6, no. 2, pp. 224–228, Apr. 2009.



Renlong Hang received the B.S. and M.S. degrees from the Nanjing University of Information Science and Technology, Nanjing, China, in 2011 and 2014, respectively, where he is currently working toward the Ph.D. degree in the School of Atmospheric Science.

His current research interests include machine learning and pattern recognition.

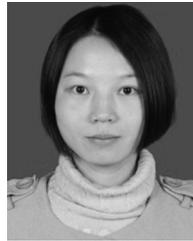


Qingshan Liu (M'05–SM'07) received the M.S. degree from Southeast University, Nanjing, China, in 2000 and the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2003.

From 2010 to 2011, he was an Assistant Research Professor with the Department of Computer Science, Computational Biomedicine Imaging and Modeling Center, Rutgers, The State University of New Jersey, Piscataway, NJ, USA. Before he joined Rutgers University, he was an Associate Professor with the National Laboratory of Pattern Recognition, Chinese

Academy of Sciences. During June 2004 and April 2005, he was an Associate Researcher with the Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong. He is currently a Professor with the School of Information and Control Engineering, Nanjing University of Information Science and Technology, Nanjing. His research interests include image and vision analysis.

Dr. Liu received the President Scholarship of the Chinese Academy of Sciences in 2003.



Huihui Song received the B.S. degree in technology and science of electronic information from the Ocean University of China, Qingdao, China, in 2008, the Masters degree in communication and information system from the University of Science and Technology of China, Hefei, China, in 2011, and the Ph.D. degree in geography and resource management from the Chinese University of Hong Kong, Hong Kong, in 2014.

She is currently a Professor with the Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing, China. Her research interests include remote sensing image processing and image fusion.

Yubao Sun received the Ph.D. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2010.

He is currently an Associate Professor with the School of Information and Control Engineering, Nanjing University of Information Science and Technology. His current research interests are sparse representation theory and application, low-rank representation, video analysis and understanding, and graph and hypergraph learning.

Fuping Zhu received the Ph.D. degree from the State Key Laboratory of Management and Control for Complex Systems, Chinese Academy of Science, Beijing, China, in 2003.

He is a Professor at Beijing Electro-Mechanical Engineering Institute, Beijing. His research interests include image processing, automatic target recognition, and image fusion.

Hucheng Pei received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998, respectively.

He is a Professor at the Beijing Electro-Mechanical Engineering Institute, Beijing, China. His research interests include signal processing and automatic target recognition.