Neural Process Lett (2016) 43:805–821
DOI 10.1007/s11063-015-9447-0

A Novel Spatial Analysis Method for Remote Sensing Image Classification

Jianqiang Gao[1] · **Lizhong Xu**[1]

Published online: 18 June 2015

Abstract A new and efficient classification model is introduced in this paper. The proposed model enjoys the information of null space of within-class and range space of within-class. And the proposed model aims at defining a reliable spatial analysis criterion for the remote sensing image, taking advantage of the differences in different areas. Finally, by incorporating fisher linear discriminant analysis and support vector machine (or K-nearest neighbor) classifier among image pixels, the model obtained more accurate classification results.

Keywords Remote sensing image · Projection matrix · Support vector machine · Image classification · Kernel method

Mathematics Subject Classification 68T10 · 68U10

1 Introduction

Remote sensing image classification has become one of the hot topic in recent twenty years that is because the image contains much useful information, which plays an important role in the development of society and economy. Remote sensing image combines imaging techniques with the spectral techniques togethers, are widely used in civil and military areas. However, remote sensing image classification is a complex process that may be affected by many factors. For instance, the availability of high quality images, proper classification approach, and the analytical ability of researchers. For a particular study, it is often difficult to identify the best classifier due to the lack of a guideline for selection and the availability of suitable classification models to band. Hence, many scientists have made great efforts

✉ Jianqiang Gao
jianqianggaohh@126.com; dr.jq.gao@gmail.com

Lizhong Xu
mathtwo@hotmail.co.in

[1] College of Computer and Information Engineering, Hohai University, Nanjing 210098, People's Republic of China

[1–10] to improve classification accuracy. In many reports, supervised, semi-supervised and unsupervised are the three popular leaning methods for remote sensing image classification, such as, maximum-likelihood classifiers, neural networks and neurofuzzy models [11–13]. However, there is an important hughes phenomenon [14] for hyper-spectral images. So, it needs a long time to deal with the high dimensional data.

The purpose of classification is to estimate the different species of each geographic region in remote sensing image. It is usually formulated as a segmentation task where an appearance model is first used to filter the pixel and then threshold setting strategies are utilized to infer the affiliation of the pixel in current frame. Hence, how to effectively model the appearance of the target region and how to accurately infer the affiliation from all ground-based ancillary data are two main steps for a successful classification system. Although a variety of classification algorithms have been proposed, remote sensing image classification still cannot meet the requirements of most practical applications. Many elegant features in the field of pattern recognition can be used to discriminate the category from image and ancillary data. However, the extraction of useful information and modeling are very difficult to achieve because the remote sensing image exhibits more complex spectral character, high-dimensional data, and the selection of band. Hence, traditional models to achieve robust classification are not always feasible.

In recent years, Quan et al. [15] proposed a multiscale segmentation method, which is combining the probabilistic neural network (PNN) with the multiscale autoregressive model. A hybird classifier was proposed by Zhang et al. [16] for polarimetric synthetic aperture radar (SAR) images. Besides, graph-based learning method is becoming an emergent research topic in image classification, some related research can be found in [17–20]. In addition, multiview learning approach based image annotation also becoming one of the hot topic in image processing [21,22]. However, for hyper-spectral remote image classification, the challenge is to develop approaches that are powerful enough to make use of the intricate details. While the growing number of spectral channels enables discrimination among a large number of cover classes, lots of traditional algorithms fail on these data because of mathematical or practical limitations. For instance, the maximum likelihood and other covariance-based classifiers require, as many training samples each class as the number of bands plus one, which produces a severe problem of field sampling for multi-bands hyper-spectral image. In high dimensional data analysis, such as hyper-spectral remote sensing image, the dimension reduction process plays an important role in all supervised or unsupervised classification approaches which require the estimation of second order statistics. The aim of dimension reduction process is to map a set of high dimensional data into a low dimensional space while preserving the intrinsic structure of that data. The most famous dimension reduction methods have been reported in the literature, such as principal component analysis (PCA) and its generalized kernel PCA [23,24], locally linear embedding (LLE) [25] and spatial and spectral oriented dimension reduction (SASO-DR) [26]. However, this operation of dimension reduction can result in an undesirable loss of information. In addition, many researchers believe that the kernel trick is one of the best tools for solving high dimensional data classification [27–30] or pattern recognition problem [31–33]. Actually, the kernel trick [34] is applied to project the original data into a feature space in which the data become linearly (or approximate linearly) separable. Fauvel et al. [27] proposed a spatial-spectral kernel-based approach with the spatial and spectral information were jointly used for the classification. A kernel-based block matrix decomposition approach for the remote sensing image classification can be found in [35].

The main goal of this paper is to establish a new and efficient classification model applicable in remote sensing image processing by using mathematical theory. The model

is a three-step process with adjustment of parameters. (1) The information of same areas associated with each pixel is modeled as the within-class set which generates within-class scatter matrix **Sw**, at the same time the information of different areas associated with mean pixel of each same areas is modeled as the between-class set which generates between-class scatter matrix **Sb**. (2) A projection matrix **W** (**W** = [$\alpha \cdot$**W1**, $(1-\alpha)\cdot$**W2**]) can be obtained by solving an optimal problem with using fisher linear discriminant analysis (FLDA) approach in the null space of **Sw** and range space of **Sw**. Where α $(0 \leq \alpha \leq 1)$ is a non-negative parameter constant to balance the null space of **Sw** and range space of **Sw**. (3) The projection matrix **W** was used to project the original data into a new low-dimension feature space, and then support vector machine (SVM) (or KNN) classifier was used in the last step. The advantage of the proposed model is that the spatial information of remote sensing image can be fully used in classification processing. So, we abbreviate that by calling it parameterized null-range-**Sw**.

The remainder of this paper is organized as follows. Section 2 briefly reviews the formulations of FLDA, KNN and SVM. In Sect. 3, the derivation process of the proposed model is described in detail. The effectiveness of the model is demonstrated in Sect. 4 by experiments on several real remotely sensed images. Finally, Sect. 5 concludes this paper.

2 Review of FLDA, KNN and SVM

2.1 Fisher Linear Discriminant Analysis (FLDA)

The main idea of FLDA is to perform dimension reduction while preserving as much information as possible. Linear discriminant analysis (LDA) aims to find the optimal projection matrix such that the class structure of the original high-dimensional space is preserved in the low-dimensional space. However, the FLDA cannot be directly applied because the **Sw** has zero eigenvalues. Hence, many methods are proposed to solve it, such as LDA/QR [36], null subspace method [37], range subspace method [38] and based-median method [39,40].

In this subsection, we first introduce some important notations used in this paper. Let c be the number of classes, N_i be the number of samples from ith class, N be the number of total samples from each class, A_j^i be the jth sample from ith class and m_i be the mean of ith class samples.

$$N = \sum_{i=1}^{c} N_i, \tag{1}$$

$$m_i = \frac{1}{N} \sum_{j=1}^{N_i} A_j^i, \quad (i = 1, \ldots, c). \tag{2}$$

The optimal projection matrix $W = [w_1, w_2, \ldots, w_r]$ can be obtained via maximizing the following criterion [41]. Where r is at most $\min(c-1,\ N)$.

$$J(W) = \frac{W^T \mathbf{Sb} W}{W^T \mathbf{Sw} W}, \tag{3}$$

where, **Sb** and **Sw** are the between-class and within-class scatter matrices, respectively. m_0 is the global mean of all classes samples.

$$\mathbf{Sb} = \sum_{i=1}^{c} (m_i - m_0)^T (m_i - m_0), \tag{4}$$

$$\mathbf{Sw} = \sum_{i=1}^{c} \sum_{j=1}^{N_i} (A_j^i - m_i)^T (A_j^i - m_i), \tag{5}$$

$$m_0 = \frac{1}{c} \sum_{i=1}^{c} m_i. \tag{6}$$

Pan et al. [42] gave a spectral regression discriminant analysis for hyperspectral image classification. Ghosh et al. [43] proposed a context-sensitive technique for unsupervised change detection in multitemporal remote sensing images. Bandos et al. [44] analyzed the classification of hyperspectral remote sensing image with LDA in the presence of a small ratio between the number of training samples and the number of spectral features.

2.2 K-Nearest Neighbor (KNN)

K-nearest neighbor is a nonparametric approach for classification. It does not require the priori knowledge such as priori probabilities and the conditional probabilities. It operates directly towards the samples and is categorized as an instance-based classification method. Details can be found in the [45,46].

2.3 Support Vector Machine (SVM)

In this subsection, we briefly review the support vector machine. Given a labeled training set $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, +1\}$, and $\Phi(\cdot)$ is a nonlinear mapping. Generally speaking, to a high-dimension space, $\Phi : \mathbb{R}^d \to \mathbb{H}$, the SVM algorithm solves

$$\min_{w,\xi_i,b} \left\{ \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \right\}, \tag{7}$$

constrained to

$$y_i(\langle \Phi(x_i), w \rangle + b) \geq 1 - \xi_i, \quad \forall i = 1, 2, \ldots, N. \tag{8}$$

$$\xi_i \geq 0, \quad \forall i = 1, 2, \ldots, N. \tag{9}$$

where w and b define a linear classifier in the feature space. According to the Cover's theorem [47], the nonlinear mapping function Φ is used in the transformed samples feature space. The parameter C controls the generalization capabilities of the classifier and it must be selected by the user, and ξ_i are positive slack variables enabling to deal with permitted errors.

In mathematics, the primal problem (7) is solved through its Lagrangian dual problem (10) because of the high-dimension of vector w.

$$\max_{\alpha_i} \left\{ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \Phi(x_i), \Phi(x_j) \rangle \right\}, \tag{10}$$

constrained to $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i y_i = 0, i = 1, 2, \ldots, N$. Where auxiliary variables α_i are Lagrange multipliers corresponding to constraints in (8). All Φ mappings are performed in the form of inner products. So, a kernel function K needs to be defined as (11).

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle. \tag{11}$$

By introducing (11) into (10), and then solving the dual problem, we can obtain the solution $w = \sum_{i=1}^{N} y_i \alpha_i \Phi(x_i)$. For any test vector x, the decision function can be obtained as below.

$$f(x) = sgn\left(\sum_{i=1}^{N} y_i \alpha_i K(x_i, x) + b \right), \tag{12}$$

where b can be easily obtained from the α_i that are neither 0 nor C. Details can be found in [48]. The SVM technique is widely used in remote sensing image can be found in [27–29,49–52].

3 Proposed Parameterized Null-Range-Sw (P-NRSw) Model

3.1 P-NRSw Model

In this subsection, our motivation came from the reserve of several details of remote sensing image. In our proposed model, we treat the same class areas and different class areas of remote sensing image as the within-class scatter matrix (**Sw**) and between-class scatter matrix (**Sb**), respectively. To distinguish each area of remote sensing image, we hope the differences as small as possible from the same areas, on the contrary, the differences as large as possible from the different areas. Meanwhile, inspired and motivated by the idea of FLDA, we proposed P-NRSw model to deal with the same and different areas of remote sensing image.

From a mathematical theory point of view, the proposed P-NRSw model consists of three subsequent steps. Firstly, the P-NRSw model projects the original space onto the null space of **Sw** using an orthogonal basis of null (**Sw**), and then in the projected space, a transformation that maximizes the between-class scatter is computed (generate project matrix **W1**). At the same time, the P-NRSw model projects the original space onto the range space of **Sw** by using a basis of range (**Sw**), and then in the transformed space the maximization of between-class scatter is pursued (generate project matrix **W2**). Secondly, select the appropriate parameters α $(0 \leq \alpha \leq 1)$ while the project matrix is constructed as $\mathbf{W} = [\alpha \cdot \mathbf{W1}, (1 - \alpha) \cdot \mathbf{W2}]$. Where α is a non-negative parameter constant to balance the null space of **Sw** and range space of **Sw**. Finally, the projection matrix **W** was used to project the original data into a new low-dimension feature space, and then SVM (or KNN) classifier was used in the last step. The details can be found in the followings:

In the following, first a brief description of some important notations are given in this section. For convenience, we list these notations in Table 1.

Consider a original multi-band remote sensing image with N samples (different class areas) where each sample contains d bands. By this assumption, a $N \times d$ matrix X can be constructed which has the components x_1 to x_N, where each component is consists of the bands of remote sensing image. Therefore, the dataset that consists of N samples $\{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$, and $y_i \in \{1, 2, \ldots, c\}$ denotes the class label of the ith sample, N is the sample size, d is the data dimensionality, and c is the number of classes. Let the data matrix $X = [x_1, x_2, \ldots, x_N]$ be partitioned into c classes as $X = [X_1, X_2, \ldots, X_c]$, where

Table 1 Notation description

Notation	Description
X	Data matrix
X_i	ith class data matrix
N	The number of samples
H_b	Precursor of between-class scatter
H_w	Precursor of within-class scatter
Sb	Between-class scatter matrix
Sw	Within-class scatter matrix
W1	Projection matrix form null space **Sw**
W2	Projection matrix form range space **Sw**
W	Combination projection matrix form **W1** and **W2** With parameter α
α	Non-negative parameter

$X_i \in \mathbb{R}^{d\times n_i}$, n_i is the size of the ith class X_i and $\sum_{i=1}^{c} n_i = N$. Define matrices H_w and H_b as follows:

$$H_w = \frac{1}{\sqrt{N}}\left[X_1 - m_1 e_1^T, \ldots, X_c - m_c e_c^T\right], \tag{13}$$

$$H_b = \frac{1}{\sqrt{N}}\left[\sqrt{n_1}(m_1 - m_0), \ldots, \sqrt{n_c}(m_c - m_0)\right], \tag{14}$$

where m_i is the centroid of the ith class, m_0 is the global centroid, e_i is the vector of all ones of length n_i. $H_w \in \mathbb{R}^{d\times N}$, $H_b \in \mathbb{R}^{d\times c}$. Then **Sw** and **Sb** can be expressed as follows:

$$\mathbf{Sw} = H_w H_w^T, \tag{15}$$

$$\mathbf{Sb} = H_b H_b^T. \tag{16}$$

Chen et al. [37] finds solution vectors in null (**Sw**) and Yu and Yang [38] restricts the space to range (**Sb**). However, these methods may ignore this fact that the solution vectors may from two different spaces. Hence, the proposed P-NRSw model to obtain solution vectors in both spaces. Considering the singular value decomposition (SVD) of $\mathbf{Sw} \in \mathbb{R}^{d\times d}$.

$$\mathbf{Sw} = U_w \Sigma_w U_w^T = [\underbrace{U_{w1}}_{s}\underbrace{U_{w2}}_{d-s}]\begin{bmatrix}\Sigma_{w1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}U_{w1}^T\\ U_{w2}^T\end{bmatrix}, \tag{17}$$

where s =rank (**Sw**), null (**Sw**) = span(U_{w2}), $U_w \in \mathbb{R}^{d\times d}$, $U_{w1} \in \mathbb{R}^{d\times s}$, $U_{w2} \in \mathbb{R}^{d\times(d-s)}$. In the transformed space by U_{w2}, let the between-class scatter matrix be $\widetilde{S}_b = U_{w2}^T \mathbf{Sb} U_{w2}$, $\widetilde{S}_b \in \mathbb{R}^{(d-s)\times(d-s)}$. Then the basis of range ($\widetilde{S}_b$) can be found by the SVD of $\widetilde{S}_b$ as (18).

$$\widetilde{S}_b = \widetilde{U}_b \widetilde{\Sigma}_b \widetilde{U}_b^T = [\underbrace{\widetilde{U}_{b1}}_{r1}\ \underbrace{\widetilde{U}_{b2}}_{d-s-r1}]\begin{bmatrix}\widetilde{\Sigma}_{b1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\widetilde{U}_{b1}^T\\ \widetilde{U}_{b2}^T\end{bmatrix}, \tag{18}$$

where $r1$ =rank ($\widetilde{S}_b$), $\widetilde{U}_b \in \mathbb{R}^{(d-s)\times(d-s)}$, $\widetilde{U}_{b1} \in \mathbb{R}^{(d-s)\times r1}$, $\widetilde{U}_{b2} \in \mathbb{R}^{(d-s)\times(d-s-r1)}$. In the transformed space by the basis $\widetilde{U}_{b1}$ of range ($\widetilde{S}_b$), let Y be the matrix whose columns are the eigenvectors corresponding to nonzero eigenvalues of

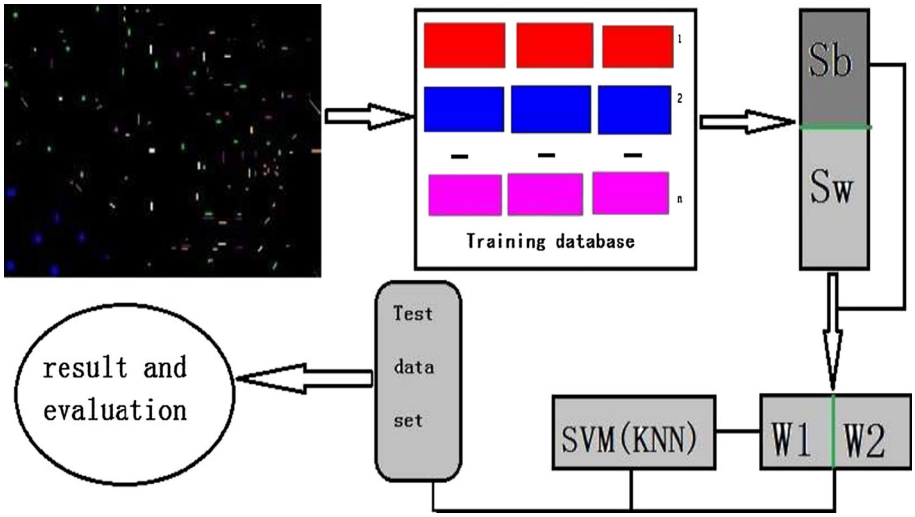

Fig. 1 Flowchart of our approach

$$S_b^* = \widetilde{U}_{b1}^T U_{w2}^T \mathbf{Sb} U_{w2} \widetilde{U}_{b1}. \tag{19}$$

On the other hand, in the transformed space by the basis U_{w1}, let the between-class scatter matrix $\widehat{S}_b = U_{w1}^T \mathbf{Sb} U_{w1}$, $\widehat{S}_b \in \mathbb{R}^{s\times s}$. Then the basis of range $(\widehat{S}_b)$ can be found by the SVD of $\widehat{S}_b$ as (20).

$$\widehat{S}_b = \widehat{U}_b \widehat{\Sigma}_b \widehat{U}_b^T = [\underbrace{\widehat{U}_{b1}}_{r2} \underbrace{\widehat{U}_{b2}}_{s-r2}] \begin{bmatrix} \widehat{\Sigma}_{b1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \widehat{U}_{b1}^T \\ \widehat{U}_{b2}^T \end{bmatrix}, \tag{20}$$

where $r2 = \text{rank}(\widehat{S}_b)$, $\widehat{U}_b \in \mathbb{R}^{s\times s}$, $\widehat{U}_{b1} \in \mathbb{R}^{s\times r2}$, $\widehat{U}_{b2} \in \mathbb{R}^{s\times(s-r2)}$. In the transformed space by the basis $\widehat{U}_{b1}$ of range $(\widehat{S}_b)$, let Z be the matrix whose columns are the eigenvectors corresponding to nonzero eigenvalues of

$$S_b^\star = \widehat{U}_{b1}^T U_{w1}^T \mathbf{Sb} U_{w1} \widehat{U}_{b1}. \tag{21}$$

Therefore, the two projection matrices $\mathbf{W1} = U_{w2}\widetilde{U}_{b1}Y$ and $\mathbf{W2} = U_{w1}\widehat{U}_{b1}Z$, which are from the null space of **Sw** and the range space of **Sw**, respectively. In order to obtain stronger discriminant information, the parameter α is introduced to **W1** and **W2** in proposed model (Viz. $\mathbf{W} = [\alpha \cdot \mathbf{W1}, (1-\alpha)\cdot \mathbf{W2}]$). Finally, the classified results of remote sensing image test data set will be obtained through **W** and SVM (or KNN). Figure 1 shows a block diagram of our method.

3.2 Analysis of Computational Complexities

In this subsection, we analyze computational complexities for the discussed methods. The computational complexity for the SVD decomposition depends on what parts need to be explicitly computed. We use flop counts for the analysis of computational complexities where one flop (floating point operation) represents roughly what is required to do one addition (subtraction) or one multiplication (division) [53]. For the SVD of a matrix $H \in \mathbb{R}^{p\times q}$ when

Table 2 Complexities description

Need to be computed explicitly	Complexities
U, Σ	$4p^2q + 13q^3$
U_1, Σ	$6pq^2 + 11q^3$
U, Σ, V	$4p^2q + 22q^3$

$p \gg q$, $H = U\Sigma V^T = [\underbrace{U_1}_{q} \underbrace{U_2}_{p-q}]\Sigma V^T$, where $U \in \mathbb{R}^{p\times p}$, $\Sigma \in \mathbb{R}^{p\times q}$ and $V \in \mathbb{R}^{q\times q}$, the complexities (flops) can be roughly estimated as follows (Table 2).

For the multiplication of the $p1 \times p2$ matrix and the $p2 \times p3$ matrix, $2p1p2p3$ flops can be counted. Compared to the traditional FLDA, null subspace approach [37], range subspace method [38] and popular morphology-based feature extraction method, which involves careful design and complex steps, such as, a series of opening and closing operations, our proposed method is more efficient and computationally straightforward. Moreover, with a SVD decomposition and a simple sorting process of eigenvectors, while the morphology-based approach cannot achieve it. As a result, the proposed P-NRSw is more discriminative for remote sensing image classification, which is confirmed by our experiments in the following section.

4 Experimental Results and Analysis

In this section, we demonstrate the effectiveness of the proposed P-NRSw model on remote sensing image classification tasks. The two publicly available databases namely San Francisco of 2012 GRSS data fusion contest (2012 GRSS data) and KSC-AVIRIS data, respectively.

It is well known that the selection of kernel function is very important to achieve better performance in classification tasks. Polynomial kernel [PK, Eq. (22)], Gaussian kernel [GK, Eq. (23)], and Sigmoid kernel [SK, Eq. (24)] are three commonly used kernel functions. To evaluate the efficiency of proposed model, the three kernel functions are used in our experiment.

$$k(x, y) = (g \cdot x \cdot y + 1)^3. \tag{22}$$

$$k(x, y) = \exp(g \cdot ||x - y||^2). \tag{23}$$

$$k(x, y) = \tanh(g \cdot x \cdot y - 1). \tag{24}$$

In SVM, the penalty term C and the width of kernel g are need to be tuned. And the libsvm [54] was used. Each original remote sensing image dataset was scaled between [0, 1] by using a per band range stretching method. The within classification accuracy (WCA) and the total classification accuracy (TCA) criteria [26] are used in our comparison.

$$WCAi = \frac{P_i}{M_i} \times 100, \tag{25}$$

$$TCA = \frac{\sum_{i=1}^{c} P_i}{P} \times 100. \tag{26}$$

In these equations, P_i denotes the number of correctly classified samples in ith class, M_i is the number of samples in ith class, c is the number of classes, and P is the total number

Fig. 2 The images of training set (**a**) and test set (**b**)

Table 3 Number of the training and test samples of 2012 GRSS data

(ith class) Information classes	No. of training samples	No. of test samples
(1) Grass	200	6200
(2) Road	200	1601
(3) Roof	200	2605
(4) Shadow	200	1427
(5) Trail	200	1142
(6) Tree	200	2223
(7) Water	200	5100
Total	1400	20,298

samples in the test data. In addition, the kappa coefficient (κ) was calculated using confusion matrix in our experiments.

4.1 2012 GRSS Data

The data fusion contest has been annually organized by the data fusion technical committee of the IEEE Geoscience and Remote Sensing Society (GRSS) since 2006. Detailed description can be found in [55]. In our experiments, the data set is a subset of 477×342 pixels of original image. Meanwhile, we only considered seven classes such as grass, road, roof, shadow, trail, tree, and water, to characterize this area. The distributions of training set and test set are shown in Fig. 2. The class definitions and the number of samples for each experiment is listed in Table 3.

In SVM, the parameters C ($1 \leq C \leq 10$) and g ($2^{-10} \leq g \leq 2^{10}$)are determined by five fold cross validation strategy. Performance indicators (PIS) denotes WCAi ($i = 1, \ldots, c$), TCA, and κ. The results by using the above mentioned three kernels ($\alpha = 0.5$) are reported in Tables 4, 5 and 6, respectively.

According to the results in Tables 4, 5 and 6, the proposed P-NRSw method gives a slightly better results in terms of total classification accuracy and kappa value. No matter what parameters C we use, the P-NRSw almost give a slightly better results. When the sigmoid kernel was used in this experiment, four methods give a worse results. So, the kernel function is very important to achieve better classification. From a practical point of view, the

Table 4 TCA (%) and κ of 2012 GRSS data (Polynomial kernel)

Method	$C = 2$	$g = 223$	$C = 7$	$g = 223$
	TCA	κ	TCA	κ
N(Sw)	87.1416	0.8399	87.1761	0.8403
R(Sb)	77.4214	0.7182	86.0479	0.8265
SVM	87.1527	0.8400	87.0500	0.8411
P-NRSw	87.2549	0.8412	87.5209	0.8446

Table 5 TCA (%) and κ of 2012 GRSS data (Gaussian kernel)

Method	$C = 25$	$g = 240$	$C = 30$	$g = 240$
	TCA	κ	TCA	κ
N(Sw)	87.6293	0.8459	87.6293	0.8459
R(Sb)	86.0922	0.8271	86.0479	0.8265
SVM	88.0310	0.8510	88.1401	0.8522
P-NRSw	88.1959	0.8529	88.2304	0.8534

Table 6 TCA (%) and κ of 2012 GRSS data (Sigmoid kernel)

Method	$C = 1100$	$g = 0.05$	$C = 1200$	$g = 0.05$
	TCA	κ	TCA	κ
N(Sw)	80.4168	0.7513	81.4908	0.7661
R(Sb)	76.8007	0.7122	76.8302	0.7126
SVM	82.0584	0.7825	83.9015	0.7969
P-NRSw	83.9097	0.7972	84.4221	0.8038

gaussian kernel is the best choose. Figure 3 shows the classification results by using KNN classifier with different k values.

According to the results in Fig. 3, the proposed P-NRSw achieves the best classification results in terms of total classification accuracy. Compared with SVM, the KNN classifier is the best for 2012 GRSS data. In addition, the influence of the parameter α will be investigated in the following experiment, which is shown in Fig. 4. Figure 4 shows that the TCA rises at first and then keeps stable and then it begins descend slowly; but when the value of $\alpha = 0.5$, the TCA is the best.

In order to further analyze the effectiveness of P-NRSw model, the KSC-AVIRIS data is used for the next experiment.

4.2 KSC-AVIRIS Data

Detailed description of KSC-AVIRIS data can be found in [35,55]. The image is 614×512 pixels. Fig. 5 shows the original image with brands 11, 21, and 31. In this data set, there are lots of singular points. So, In the experiments, the singular point was replaced by using its surrounding mean value. 50 samples are randomly taken from each class as training samples

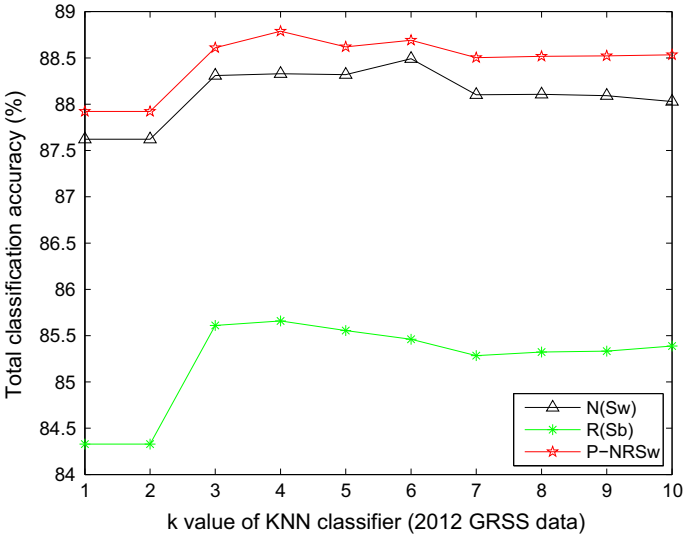

Fig. 3 TCA of different k values (2012 GRSS data, $\alpha = 0.5$)

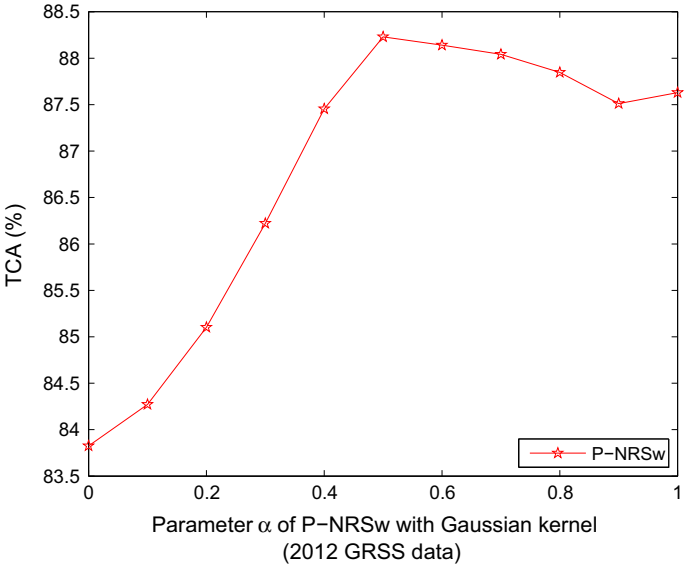

Fig. 4 TCA of different parameters α of P-NRSw with Gaussian kernel (2012 GRSS data)

and the rest are used test samples (see Table 7). In order to evaluate the generalization power of method more accurately, we adopt five fold across validation strategy.

In SVM, the parameters C and g (from 2^{-10} to 2^{10}, the step is $2^{0.2}$) are determined by five fold cross validation strategy. In the following experiments, the $\alpha = 0.5$. The optimal parameter and results are shown in Tables 8, 9 and 10 with different kernels.

From Tables 8, 9 and 10, the proposed P-NRSw performs better than other three methods in terms of total classification accuracies and kappa coefficient. In addition, according to

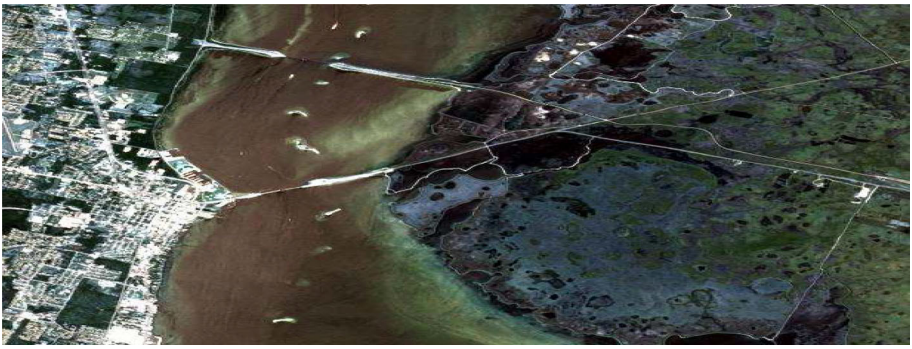

Fig. 5 KSC-AVIRIS data, (Bands 11, 21, 31) acquired over KSC

Table 7 Class codes, names, and number of the training and test samples (KSC-AVIRIS)

(ith class) Information classes	No. samples	No. of training samples	No. of test samples
(1) Scrub	761	50	711
(2) Willow swamp	243	50	193
(3) Cabbage palm hammock	256	50	206
(4) Cabbage palm/oak	252	50	202
(5) Slash pine	161	50	111
(6) Oak/broadleaf hammock	229	50	179
(7) Hardwood swamp	105	50	55
(8) Graminoid marsh	431	50	381
(9) Spartina marsh	520	50	470
(10) Cattail marsh	404	50	354
(11) Salt marsh	419	50	369
(12) Mud flats	503	50	453
(13) Water	927	50	877
Total	5211	650	4561

Table 8 TCA (%) and κ of KSC-AVIRIS data (Polynomial kernel)

Method	$C = 2^{3.4}$ $g = 2^{0.6}$ N(Sw)	$C = 2^{7.8}$ $g = 2^{-9.6}$ R(Sb)	$C = 2^{4.4}$ $g = 2^{-5.8}$ SVM	$C = 2^{7.4}$ $g = 2^{-1.2}$ P-NRSw
TCA	91.5589	92.8744	92.0000	92.9840
κ	0.9055	0.9202	0.9102	0.9214

the TCA and α, we also found that the best result (TCA = 93.1155%, κ = 0.9229) is from gaussian kernel. Figure 6 shows the classification results by using KNN classifier with different k values.

According to Fig. 6, no matter what k we use, the P-NRSw almost gives a better result. Therefore, the robustness is significant. However, compared with SVM, the result is not ideal

Table 9 TCA (%) and κ of KSC-AVIRIS data (Gaussian kernel)

Method	$C = 2^{0.8}$ $g = 2^{6.2}$ N(Sw)	$C = 2^{9}$ $g = 2^{-10}$ R(Sb)	$C = 2^{6.6}$ $g = 2^{-7.4}$ SVM	$C = 2^{8}$ $g = 2^{0.8}$ P-NRSw
TCA	91.6027	92.9182	90.7825	93.1155
κ	0.9060	0.9207	0.8997	0.9229

Table 10 TCA (%) and κ of KSC-AVIRIS data (Sigmoid kernel)

Method	$C = 2^{5}$ $g = 2^{1.6}$ N(Sw)	$C = 2^{8.8}$ $g = 2^{-7.8}$ R(Sb)	$C = 2^{10}$ $g = 2^{-7.8}$ SVM	$C = 2^{8.6}$ $g = 2^{0.6}$ P-NRSw
TCA	91.5589	92.5016	90.8100	92.6990
κ	0.9055	0.9161	0.8970	0.9182

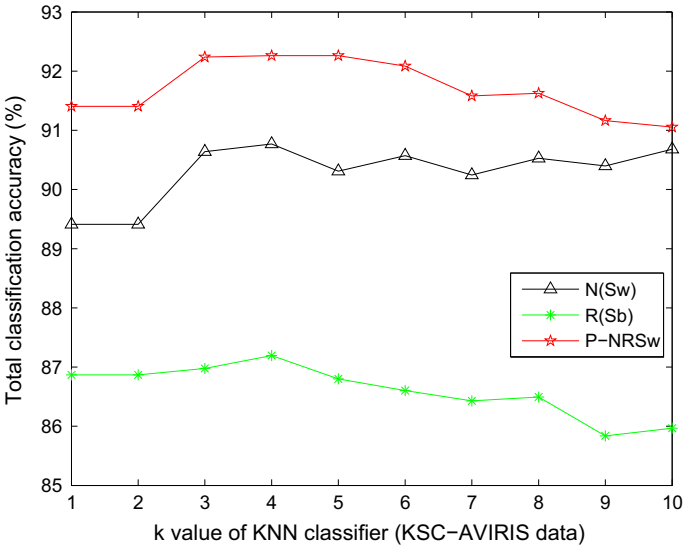

Fig. 6 TCA of different k values (KSC-AVIRIS data, $\alpha = 0.5$)

by using KNN classifier for KSC-AVIRIS data. Of course, we also found that the classification result is the best with k = 4. Table 11 shows the classification result of each class.

The results of Table 11 clearly show the superiority of the proposed method. Meanwhile, for the "Willow swamp", "Cabbage palm hammock", "Cabbage palm/oak", "Graminoid marsh", "Spartina marsh", "Cattail marsh", "Salt marsh", and "Mud flats" classes the P-NRSw produces better classification results.

In addition, according to Tables 4, 5, 6, 8, 9 and 10, we can see that P-NRSw shows better performance as compared to N(Sw), R(Sb), and SVM in terms of TCA and κ. As reported before, the classification accuracies achieved via P-NRSw are higher. So, the parameter α plays an important role in classification by contacting with the null space of **Sw** and the range space of **Sw**. Hence, the influence of the parameter α also will be investigated in the following experiment, which is shown in Fig. 7, which is displays the results of TCA with

Table 11 Classification accuracies (%) and κ of KSC-AVIRIS data (KNN, k=4)

PIS	N(Sw)	R(Sb)	P-NRSw
WCA1	84.3882	81.1533	89.7328
WCA2	93.2642	86.0104	**93.7824**
WCA3	83.0097	75.2427	**88.3495**
WCA4	57.9208	58.9109	**65.8416**
WCA5	84.6847	73.8739	82.8829
WCA6	67.5978	58.6592	62.0112
WCA7	92.7273	72.7273	87.2727
WCA8	90.8136	86.8766	**91.0761**
WCA9	98.0851	92.1277	**98.9362**
WCA10	96.6102	94.9153	**97.1751**
WCA11	90.2439	89.9729	**92.6829**
WCA12	98.6755	93.5982	**98.8962**
WCA13	100	100	100
TCA	90.7696	87.1958	**92.2605**
κ	0.8966	0.8567	**0.9132**

Bold values indicate the best result for P-NRSw method

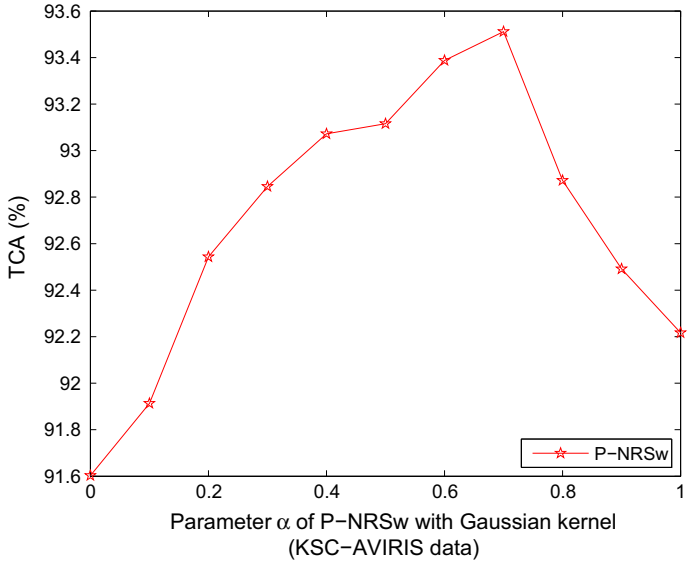

Fig. 7 TCA of different parameters α of P-NRSw with Gaussian kernel (KSC-AVIRIS data)

different parameters α. And meanwhile, it shows that the TCA rises at first and then it begins descend fast; but when the value of $\alpha = 0.7$, the TCA is the best. We know that the value of α is used to control the contribution between the null space of **Sw** and the range space of **Sw**.

Compared the classification results with Tables 4, 5, 6, 8, 9 and 10, gaussian kernel give the best result. From what has been discussed above, the P-NRSw outperforms other three

Table 12 TCA (%), κ and times (s) for N(Sw), R(Sb), SVM, P-SVM and P-NRSw of different data sets

Method	#1			#2		
	TCA	κ	Time	TCA	κ	Time
N(Sw) [37]	87.6293	0.8459	161	91.6027	0.9060	47
R(Sb) [38]	86.0922	0.8271	158	92.9182	0.9207	52
SVM [10]	88.1401	0.8522	198	92.0000	0.9102	56
P-SVM [10]	88.1585	0.8528	212	93.0241	0.9118	67
P-NRSw	88.2304	0.8534	225	93.1155	0.9229	74

methods. In addition, from a theoretical point of view, P-NRSw offers a stable method to handle remote sensing image classification problems.

In order to further verify the effectiveness of the proposed model, we made the following experiment. Table 12 gives the total classification accuracy (TCA), kappa coefficient and time cost of different data sets. #1 and #2 denote the 2012 GRSS data set and KSC-AVIRIS data set, respectively. All the methods were implemented by using MATLAB R2010b on a desktop PC equipped with an Intel Core 2 i3 (at 2.40 GHz) and 2 GB of RAM memory.

According to Table 12, we can see that P-NRSw shows better performance as compared to P-SVM, SVM, R(Sb) and N(Sw) in terms of TCA and kappa value. However, it is worth noting that the elapsed time of P-NRSw algorithm is more than other algorithms. This is because the most time consuming step is the calculation of decomposition from **Sw** and **Sb**. In addition, compared with the P-SVM method, the proposed P-NRSw method can provide higher classification accuracies with respect to TCA.

5 Conclusion and Future Work

In this paper, a new and efficient remote sensing image classification model P-NRSw been proposed. The goal of the proposed model is to make full use of detail information to search an effective projection matrix for improving the performance of image classification. For remote sensing image, the P-NRSw takes advantage of within-class similarity and between-class diversity to build a mathematical model. Then, based on the FLDA, the SVM (or KNN) is applied to obtain classification results. Experimental results obtained on two data sets confirm the effectiveness of the proposed P-NRSw, which provided efficient discriminant information for remote sensing image classification tasks. In addition, the authors realize that more work must be done to improve the classification results in the further.

Acknowledgments The authors are very grateful to the editor and anonymous referees reviews for their valuable comments and helpful suggestions. In addition, this work is supported by National Natural Science Foundation of P.R. China (Grant No. 61271386), and the Graduates' Research Innovation Program of Higher Education of Jiangsu Province of P.R. China (Grant No. CXZZ13-0239).

References

1. Gong P, Howarth PJ (1992) Frequency-based contextual classification and gray-level vector reduction for land-use identification. Photogramm Eng Remote Sens 58(4):423–437

2. Kontoes C, Wilkinson GG, Burrill A, Goffredoa S, Megier J (1993) An experimental system for the integration of GIS data in knowledge-based image analysis for remote sensing of agriculture. Int J Geogr Inf Syst 7(3):247–262
3. Foody GM (1996) Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. Int J Remote Sens 17(7):1317–1340
4. San Miguel-Ayanz J, Biging GS (1997) Comparison of single-stage and multi-stage classification approaches for cover type mapping with TM and SPOT data. Remote Sens Environ 59(1):92–104
5. Aplin P, Atkinson PM, Curran P (1999) Per-field classification of land use using the forthcoming very fine spatial resolution satellite sensors: problems and potential solutions. In: Atkinson PM, Tate NJ (eds) Advances in remote sensing and GIS analysis. Wiley, New York, pp 219–239
6. Stuckens J, Coppin PR, Bauer ME (2000) Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sens Environ 71(3):282–296
7. Franklin SE, Peddle DR, Dechka JA, Stenhouse GB (2002) Evidential reasoning with Landsat TM, DEM and GIS data for landcover classification in support of grizzly bear habitat mapping. Int J Remote Sens 23(21):4633–4652
8. Pal M, Mather PM (2003) An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens Environ 86(4):554–565
9. Gallego FJ (2004) Remote sensing and land cover area estimation. Int J Remote Sens 25(15):3019–3047
10. Zhang R, Ma J (2008) An improved SVM method P-SVM for classification of remotely sensed data. Int J Remote Sens 29(20):6029–6036
11. Melgani F, Serpico SB (2002) A statistical approach to the fusion of spectral and spatio-temporal contextual information for the classification of remote-sensing images. Pattern Recognit Lett 23(9):1053–1061
12. Bardossy A, Samaniego L (2002) Fuzzy rule-based classification of remotely sensed imagery. IEEE Trans Geosci Remote Sens 40(2):362–374
13. Bruzzone L, Cossu R (2002) A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps. IEEE Trans Geosci Remote Sens 40(9):1984–1996
14. Hughes G (1968) On the mean accuracy of statistical pattern recognizers. IEEE Trans Inf Theory 14(1):55–63
15. Quan JJ, Wen XB, Xu XQ (2008) Multiscale probabilistic neural network method for SAR image segmentation. Appl Math Comput 205(2):578–583
16. Zhang Y, Wu L, Neggaz N, Wang S, Wei G (2009) Remote-sensing image classification based on an improved probabilistic neural network. Sensors 9(9):7516–7539
17. Yu J, Tao D, Wang M (2012) Adaptive hypergraph learning and its application in image classification. IEEE Trans Image Process 21(7):3262–3272
18. Yu J, Tao D, Li J, Cheng J (2014) Semantic preserving distance metric learning and applications. Inf Sci 281:674–686
19. Yu J, Rui Y, Tao D (2014) Click prediction for web image reranking using multimodal sparse coding. IEEE Trans Image Process 23(5):2019–2032
20. Yu J, Rui Y, Tang YY, Tao D (2014) High-order distance-based multiview stochastic learning in image classification. IEEE Trans Cybern. doi:10.1109/TCYB.2014.2307862
21. Liu W, Tao D (2013) Multiview hessian regularization for image annotation. IEEE Trans Image Process 22(7):2676–2687
22. Liu W, Tao D, Cheng J, Tang Y (2014) Multiview hessian discriminative sparse coding for image annotation. Comput Vis Image Underst 118:50–60
23. Jolliffe IT (2002) Principal component analysis. Springer, Berlin
24. Schölkopf B, Smola A, Muller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10(5):1299–1319
25. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326
26. Dianat R, Kasaei S (2010) Dimension reduction of remote sensing images by incorporating spatial and spectral properties. AEU-Int J Electron Commun 64(8):729–732
27. Fauvel M, Chanussot J, Benediktsson JA (2012) A spatial-spectral kernel-based approach for the classification of remote-sensing images. Pattern Recognit 45(1):381–392
28. Fauvel M et al (2013) Advances in spectral-spatial classification of hyperspectral images. Proc IEEE 101(3):652–675
29. Liu J, Wu Z, Wei Z, Xiao L, Sun L (2013) Spatial-spectral kernel sparse representation for hyperspectral image classification. IEEE J Sel Top Appl Earth Obs Remote Sens 6(6):2462–2471
30. Camps-Valls G, Bruzzone L (2005) Kernel-based methods for hyperspectral image classification. IEEE Trans Geosci Remote Sens 43(6):1351–1362

31. Gao J, Fan L (2011) Kernel-based weighted discriminant analysis with QR decomposition and its application face recognition. WSEAS Trans Math 10(10):358–367
32. Gao J, Li L, Fan L, Xu L (2013) An application of weighted kernel fuzzy discriminant analysis. Adv Comput Math Appl 2(4):329–338
33. Gao J, Fan L, Li L, Xu L (2013) A practical application of kernel-based fuzzy discriminant analysis. Int J Appl Math Comput Sci 23(4):887–903
34. Gao S, Tsang IWH, Chia LT (2010) Kernel sparse representation for image classification and face recognition, Computer Vision-ECCV 2010. Springer, Berlin, pp 1–14
35. Gao J, Xu L, Shi A, Huang F (2014) A kernel-based block matrix decomposition approach for the classification of remotely sensed images. Appl Math Comput 228:531–545
36. Ye J, Li Q (2005) A two-stage linear discriminant analysis via QR-decomposition. IEEE Trans Pattern Anal Mach Intell 27(6):929–941
37. Chen L, Liao HM, Ko M, Lin J, Yu G (2000) A new LDA-based face recognition system which can solve the small sample size problem. Pattern Recognit 33(10):1713–1726
38. Yu H, Yang J (2001) A direct LDA algorithm for high-dimensional data-with application to face recognition. Pattern Recognit 34(10):2067–2070
39. Li X, Fei S, Zhang T (2009) Median MSD-based method for face recognition. Neurocomputing 72(16):3930–3934
40. Gao J, Fan L, Xu L (2013) Median null (Sw)-based method for face feature recognition. Appl Math Comput 219(12):6410–6419
41. Koç M, Barkana A (2011) A new solution to one sample problem in face recognition using FLDA. Appl Math Comput 217(24):10368–10376
42. Pan Y, Wu J, Huang H, Liu J (2012) Spectral regression discriminant analysis for hyperspectral image classification. ISPRS-Int Arch Photogramm Remote Sens Spat Inf Sci 1:503–508
43. Ghosh A, Mishra NS, Ghosh S (2011) Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. Inf Sci 181:699–715
44. Bandos TV, Bruzzone L, Camps-Valls G (2009) Classification of hyperspectral images with regularized linear discriminant analysis. IEEE Trans Geosci Remote Sens 47(3):862–873
45. Geva S, Sitte J (1991) Adaptive nearest neighbor pattern classification. IEEE Trans Neural Netw 2(2):318–322
46. Cover TM, Hart FE (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory IT–13(1):21–27
47. Cover TM (1965) Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition. IEEE Trans Electron Comput 14(3):326–334
48. Schölkopf B, Smola A (2002) Learning with kernels-support vector machines, regularization, optimization and beyond. MIT Press, Cambridge
49. Fauvel M, Benediktsson JA, Chanussot J, Sveinsson JR (2008) Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans Geosci Remote Sens 46(11):3804–3814
50. Tuia D, Camps-Valls G (2011) Urban image classification with semisupervised multiscale cluster kernels. IEEE J Sel Top Appl Earth Obs Remote Sens 4(1):65–74
51. Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing: a review. ISPRS J Photogramm Remote Sens 66(3):247–259
52. Sun C, Lam KM (2013) Multiple-kernel multiple-instance similarity features for efficient visual object detection. IEEE Trans Image Process 22(8):3050–3061
53. Golub GH, Van Loan CF (1996) Matrix computations, 3rd edn. Johns Hopkins University Press, Baltimore
54. Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machine http://www.csie.ntu.edu.tw/cjlin/libsvm
55. IEEE, GRSS data fusion technical committee (2012) http://www.grss-ieee.org/community/technical-committees/datafusion/