One-Class Classification of Remote Sensing Images Using Kernel Sparse Representation

Benqin Song, Peijun Li, *Senior Member, IEEE*, Jun Li, and Antonio Plaza, *Fellow, IEEE*

***Abstract*—Sparse representations have been widely studied in remote sensing image analysis in recent years. In this paper, we develop a novel method for one-class classification (OCC) using a kernel sparse representation model for remotely sensed imagery. Training samples taken from the target class alone are used to build a learning dictionary for the sparse representation model, which is then optimized to produce a reconstruction residual. In the proposed model, a pixel is classified as the target class if the obtained reconstruction residual for the pixel is smaller than a given threshold; otherwise, the pixel is labeled as the outlier class. To improve the data separability between the target and outliner classes, the training samples taken from the target class are mapped into a high-dimensional feature space using a kernel function to build a learning dictionary for the kernel sparse representation model. OCC is then conducted in the mapped high-dimensional feature space using the reconstruction residual threshold, following the same principle as OCC in the original feature space. The proposed OCC method is evaluated and compared with several existing OCC methods in three different case studies. The experimental results indicate that the proposed method outperforms these existing methods, particularly when using a kernel sparse representation.**

***Index Terms*—Kernel function, one-class classification (OCC), remote sensing imagery, sparse representations.**

I. Introduction

THE ABILITY to map and monitor land cover types and their dynamics for diverse applications has been enhanced by the availability and constantly increasing coverage of satellite images [1]. Image classification is one of the most commonly used methods to extract land cover information from remote sensing images and has been widely studied over the past three decades [2]–[4].

Various image classification techniques, supervised approaches in particular, have been developed with many successful case studies. Recently developed classification techniques include support vector machines (SVMs), random forests, and sparse representation-based methods [5], [6]. Although these supervised classifiers exhibit a very promising performance in terms of classification accuracy, they mainly focus on multiclass classification. Multiclass classifiers require all classes that occur in a study area to be exhaustively labeled [7]. Moreover, the goal in many cases is to optimize the classification accuracy for all land cover classes rather than for a specific class or few classes of interest [8]. However, in many applications, it is difficult to collect reference data for all land cover classes in the study area. It is also very common that in many applications, the focus is not on all land cover classes; instead, only one specific class or a few classes are of real interest [9], [10]. For example, some studies have targeted the identification of a specific class of interest, such as trees with yellow flowers [7] or fenland [8]. Furthermore, it is also the case in many studies that attention is only focused on a specific land cover change caused by abrupt events, such as burned areas of forests [11], buildings collapsed by earthquake disasters [12], or flooded areas [13]. In the context of such applications, the use of conventional multiclass classification techniques may be inappropriate, as it requires extra manpower and time to gather samples of classes that are ultimately of little interest given in the application goals [14]. Therefore, one-class classification (OCC) methods have emerged and are highly desirable in this context because only training samples from the target class (i.e., the single class of interest) are required [8].

Many OCC methods have been developed in relevant fields [15], [16]. These methods mainly comprise three categories: 1) density estimation methods; 2) boundary methods; and 3) reconstruction-based methods [17]. Density estimation methods directly estimate the probability density of target objects, in which an unknown test sample is judged to belong to the target class if an estimated value is higher than a given threshold. These methods include the Gaussian model [18], mixture of Gaussians [19], and Parzen density estimators [20]. On the other hand, the main objective of boundary methods is to obtain a stable and reliable boundary around the target class in all directions, which avoids the need to estimate the complete density of the data. The boundary methods include support vector data description (SVDD) [21], one-class SVM (OCSVM) [15], and the *k*-centers method [17]. Finally, reconstruction methods mainly rely on assumptions on the underlying data structures. These methods include the *k*-means method, self-organizing maps [19], and learning vector quantization [17]. In recent years, some new algorithms for OCC have been proposed, e.g., maximum entropy (MAXENT) [23]

Manuscript received July 05, 2015; revised November 09, 2015; accepted December 03, 2015. Date of publication January 25, 2016; date of current version March 11, 2016.

B. Song is with China Academy of Electronics and Information Technology, Beijing 100041, China.

P. Li is with the Institute of Remote Sensing and GIS, School of Earth and Space Sciences, Peking University, Beijing 100871, China (e-mail: pjli@pku.edu.cn).

J. Li is with Guangdong Provincial Key Laboratory of Urbanization and Geosimulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China.

A. Plaza is with Hyperspectral Computing Laboratory, University of Extremadura, 10003 Caceres, Spain.

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.2015.2508285

and one-class random forests (OCRFs) [24]. MAXENT estimates the Gibbs probability distribution that is proportional to the conditional probability of being a part of the target class. The OCRF is a method based on the random forest algorithm and an extra outlier generation procedure that employs classifier ensemble-based randomization principles [24].

OCC has also been studied in the remote sensing field in recent years. Some OCC methods originally developed in other fields have been evaluated and improved for remote sensing applications. Among these OCC methods, OCSVM [15] and SVDD [8] have been widely discussed. For example, SVDD was evaluated and quantitatively compared with other popular OCC methods, such as mixture of Gaussians and Parzen density estimators, in mapping fenland, a specific land cover class [8]. It was found that SVDD exhibited a better performance than other OCC methods. SVDD was also evaluated in hyperspectral image classification, where a multiclass strategy was proposed to classify a few land cover classes of interest (i.e., partially supervised classification) [24]. A semisupervised OCSVM was proposed and compared with biased SVM [15]. The MAXENT method was evaluated for the OCC task for land cover classification and was quantitatively compared with OCSVM [22]. The positive and unlabeled learning (PUL) algorithm, which was originally developed to handle binary classification problem [25], was proposed for OCC of remote sensing data in [14]. The PUL algorithm trains a classifier on positive and unlabeled data, estimates the probability that a positive training sample (i.e., target class sample) has been labeled, and generates binary predictions for unknown test samples using an adjusted threshold [14]. To assess the accuracy of the OCC classifiers, the effect of training sample size on OCC accuracy was discussed in [26]. Recently, a new measure for OCC accuracy assessment without negative data was proposed [27].

In addition to the development of new methods for OCC, there are also many successful studies in which the OCC methods have been used in various applications. For example, the SVDD-based method was used for anomaly detection in hyperspectral images, and the results showed improved performance [28]. An oil-slick detection method using OCSVM was proposed for synthetic aperture radar images, where a specific kernel was developed to perform accurate segmentation of the local sea-surface wave spectrum [29]. An unsupervised change detection method using SVDD was also proposed in [30]. A supervised change detection method using the OCSVM method was proposed and evaluated for burned area mapping, urban land expansion [16], and the extraction of damage to urban infrastructure (i.e., buildings and roads) [12], [31]. Although there are many studies in the development of new OCC methods, there is still a need for new OCC methods for remote sensing image classification.

Recently, sparse representation has become a widely popular approach in remote sensing data processing [32]. The sparse representation method generally assumes that an unknown test pixel can be linearly represented by a dictionary, i.e., a small number of elementary samples, often called atoms, chosen from an overcomplete training dictionary. However, there are also some studies that use undercomplete dictionaries for remote sensing data in cases where not enough labeled data exist to build an overcomplete dictionary [33], [34]. In these cases, image patches are usually used rather than pixels [33], [34]. Sparse representation has been applied to multiclass image classification [35], [36], target detection [37], and spectral unmixing [38]. Moreover, a kernel sparse representation method for image classification was recently proposed, where the samples are mapped into a high-dimensional feature space such that the obtained sparse vector contains more effective discriminative information, providing higher classification accuracies compared with conventional sparse representation-based classification methods [39]. Because of some unique characteristics of sparse representation, e.g., adaptiveness and competitiveness, the method achieved a better performance than state-of-the-art competitors [39].

It is also well known that kernel-based methods are a very powerful strategy for improving the classification accuracy. For example, the SVM method [40] is a representative kernel-based method that changes the distribution of samples by mapping them into a high-dimensional feature space, thus improving the separability between classes [40]. It has also been observed that the SVM classifier [3] can achieve higher classification accuracy than traditional classifiers [5]. Variants of kernel-based algorithms have also been developed to improve the classification accuracy [41]. These kernel-based methods include SVMs with composite kernels, which take into account both spatial and spectral information by simultaneously using more than one kernel [42], and transductive SVMs, which exploit both labeled and unlabeled samples to address ill-posed classification problems [43].

Thus, the main objective of this study is to propose a new OCC method that is based on an extended sparse representation method and that also adopts a kernel-based method to increase its discriminative capability. This represents a significant innovation with regards to many existing OCC methods that do not include these characteristics. The rest of this paper is organized as follows. In Section II, the proposed classifier is introduced. Section III describes the data sets used for evaluation purposes and performs an extensive experimental evaluation of the proposed approach with regards to state-of-the-art methods. Finally, Section IV concludes with some remarks and hints about plausible future research directions.

II. Sparse Representation-Based Classification

First, we briefly introduce the sparsity-based algorithm used for multiclass classification. In the sparsity model, it is assumed that the spectral signatures of pixels belonging to the same class approximately lie in a low-dimensional subspace [44]. Thus, an unknown test pixel can be sparsely represented by a few training samples (atoms) from a given dictionary, and the corresponding sparse representation vector will implicitly encode the class information [32]. The class label of a pixel is determined by the minimal residual between the pixel and its approximation from each class subdictionary [35].

Assume that we have a training dictionary, denoted by $\mathbf{A} = \{x_1, \ldots, x_n\} \in R^{l \times n}$, with n samples of l dimensions, comprising a total of c distinct classes. Further, assume that this dictionary is organized as $\mathbf{A} = [A_1, \ldots, A_c]$, where A_i holds

the samples of class i in its columns. Let y be an unknown test sample. In the sparse representation model, sample y is represented by a linear combination of atoms from a training dictionary $\mathbf{A}$ as follows:

$$\begin{aligned} y &= x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n + \epsilon \\ &= [x_1 x_2 \ldots x_n]\,[\alpha_1 \alpha_2 \ldots \alpha_n]^T + \epsilon = \boldsymbol{A}\boldsymbol{\alpha} + \epsilon \end{aligned} \quad (1)$$

where $\boldsymbol{\alpha} = \left[\boldsymbol{\alpha}_1^T, \ldots \boldsymbol{\alpha}_k^T \ldots, \boldsymbol{\alpha}_c^T\right]^T$ is an n-dimensional sparse vector (i.e., most elements of $\boldsymbol{\alpha}$ are zeros), $\boldsymbol{\alpha}_k$ is the vector of regression coefficients associated with class k ($k = 1, \ldots, c$), and $\in$ is the representation residual (error). As the feature vectors are very likely to lie in the convex cone spanned by the atoms of the respective class, we can assume that the coefficients α_i are nonnegative [36].

The sparse vector $\boldsymbol{\alpha}$ in (1) usually can be obtained using greedy algorithms such as basis pursuit (BP) [45] and orthogonal matching pursuit (OMP). Actually, the optimization problem involved in (1) is a combination of $\ell_2 - \ell_1$ problem, which is difficult to solve. By imposing the nonnegativity constraint, this problem can be solved using sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL) [46]. SUnSAL was initially proposed for the spectral unmixing problem [38], but was then used for image classification based on sparse representation because it is a very efficient algorithm for solving nonnegativity constraint $\ell_2 - \ell_1$ problems [36].

Finally, to introduce robustness with respect to representation error $\in$, the label of pixel y is then determined by the minimal total residual [35]

$$\widehat{class}\,(y) = \underset{k\in\{1,\ldots,c\}}{\operatorname{argmin}} \|y - \boldsymbol{A}_k\boldsymbol{\alpha}_k\|_2. \quad (2)$$

In the sparse representation model, the samples in the training set with spectral signatures similar to an unknown sample are more likely to be selected as atoms with high coefficient values when the training samples in the dictionary are representative enough [47]. Thus, unknown samples are likely to exhibit the same membership as the atoms with higher coefficients. In other words, the residuals obtained by reconstructing unknown samples using training samples from the class of higher coefficient atoms are smaller.

III. Proposed Methodology

A. OCC Using Sparse Representation

As mentioned previously, in multiclass classification using sparse representation, the class membership of an unknown sample is determined by the minimum residual between the sample and its approximation from each class subdictionary. Given that only samples of the target class (or class of interest) are available in the OCC model, and that only one residual for the target class is available, the membership of an unknown sample cannot be directly determined using the minimum residual, as in multiclass classification using traditional sparse representation [32]. Thus, a key problem in sparse representation-based OCC is how to use the residual from the target class to determine if an unknown pixel belongs to the target class instead of using the minimum residual to determine the class label of a pixel in multiclass classification. It is reasonably assumed that the smaller the residual value obtained by reconstructing a unknown sample using all training samples of the target class, the higher the probability that the unknown sample belongs to the target class. Thus, the class membership of the unknown sample can be determined using a residual threshold. By using such a threshold, the sparse representation of multiclass classification can be extended to the OCC task. Note that determining the residual threshold for a given variable is a very common strategy for many OCC tasks [17]. In this paper, we propose a sparse representation-based method for OCC, called hereinafter OCC-SR. The method is summarized as follows. Let us consider a sample set $\boldsymbol{B} = \{x_1, \ldots, x_m\} \in R^{l\times m}$, with m samples of l dimensions, belonging to a given class of interest (target class). Suppose that y is an unknown sample. First, the unknown sample y is sparsely represented using all training samples $\boldsymbol{B}$. We then seek a sparse reconstruction coefficient vector $\boldsymbol{\alpha}$ for y using the following minimization problem:

$$\begin{aligned} y &\approx x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m \\ &= [x_1 x_2 \ldots x_m]\,[\alpha_1 \alpha_2 \ldots \alpha_m]^T = \boldsymbol{B}\boldsymbol{\alpha}. \end{aligned} \quad (3)$$

The sparse representation coefficient vector is obtained when (3) is solved. In this case, the reconstruction residual r_y of y is

$$r_y = \|y - \boldsymbol{B}\boldsymbol{\alpha}\|_2. \quad (4)$$

It is reasonable that a smaller reconstruction residual for a test sample indicates that the test sample is more similar to the training samples. Conversely, a larger residual value shows a more significant spectral signature difference between the test sample and training samples. To assign the class label to a given pixel, a residual threshold T_r needs to be first determined, as in many OCC tasks [17]. For unknown sample y, if $r_y \le T_r$, y is classified as the target class; otherwise, it is classified as the outlier class.

In order to facilitate the determination of an appropriate residual threshold T_r, we propose to estimate a threshold range from the training samples of the target class. Each sample x_i from the training sample set (training dictionary) is sparsely represented using all remaining training samples and a reconstruction residual is obtained for the sample

$$r_i = \|x_i - \boldsymbol{B}_i\boldsymbol{\alpha}_i\|_2 \quad (5)$$

where $\boldsymbol{B}_i = \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_m\}$ and $\boldsymbol{\alpha}_i = \{\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots \alpha_m\}$. After all the samples in the training dictionary have been reconstructed, a residual vector $\boldsymbol{r} = [r_1, \ldots, r_m]$ is obtained. The minimum and maximum of all reconstruction residuals for all training samples are also obtained. We assume that the appropriate residual threshold T_r is located between these minimum and maximum values. To standardize the threshold selection, we adopt an acceptance rate parameter λ, which is linearly related to the residual threshold T_r

$$T_r = min\,(\mathbf{r}) + \lambda\,(max\,(\mathbf{r}) - min\,(\mathbf{r})) \quad (6)$$

where the acceptance rate parameter λ lies in the range: $0 \le \lambda \le 1$, *min*($\boldsymbol{r}$), and *max*($\boldsymbol{r}$) are the minimum and maximum

values of the reconstruction residuals estimated from training samples. Note that λ is an important parameter in the proposed method. From (6), it is clear that as λ increases, T_{r} increases. As a result, more unknown samples will be classified as belonging to the target class, and the commission error of the target class may also increase with higher λ values. Thus, the use of λ mainly helps us to determine the appropriate threshold between the minimum and maximum values of $\boldsymbol{r}$.

B. OCC Using Kernel Sparse Representation

As it is well known for kernel-based approaches (e.g., SVM) [40], after mapping samples into a high-dimensional feature space by a nonlinear mapping function (i.e., kernel function), the distribution of samples can be changed [15]. This change may make linearly inseparable samples in the original spectral feature space become linearly separable in the mapped high-dimensional feature space [39]. In [39], the kernel sparse representation, which maps all the samples into a high-dimensional feature space, was proposed for multiclass classification and showed more powerful discrimination ability than the original sparse representation-based classification. Inspired by this, kernel sparse representation was extended to OCC in this study. Hereafter, we refer to this method as OCC using the kernel sparse representation (OCC-KSR). In the proposed OCC-KSR, all samples are first mapped into a high-dimensional feature space and then OCC-SR is performed in this space.

Suppose that ϕ is a nonlinear mapping function (i.e., a kernel function). A training sample x_i mapped from the original spectral feature space R^l into a high-dimensional feature space H can be expressed as $\phi(x_i)$. Thus, $\boldsymbol{C}_i = [\phi(x_1), \ldots, \phi(x_{i-1}), \phi(x_{i+1}), \ldots, \phi(x_m)]$ represents all remaining samples in the training dictionary in the mapped high-dimensional feature space. The sparse representation problem in H can be described as [40]

$$\begin{aligned}\phi(x_i) \approx{}& \phi(x_1)\beta_1 \cdots + \phi(x_{i-1})\beta_{\mathrm{i-1}} \\ &+ \phi(x_{i+1})\beta_{\mathrm{i+1}} + \cdots + \phi(x_m)\beta_m = \boldsymbol{C_i}\boldsymbol{\beta_i} \quad (7)\end{aligned}$$

where x_i is a training sample in the original spectral space, and $\phi(x_i)$ is the sample mapped into the high-dimensional feature space H. $\boldsymbol{\beta_i}$ is the sparse vector in H. Equation (7) is similar to (1). The difference between two equations is that the former is in the high-dimensional feature space and the latter is in the original spectral space. Sparse vector $\boldsymbol{\beta_i}$ in H can be obtained as

$$\min_{\beta_i} \|\boldsymbol{\beta_i}\|_0 \text{ subject to } \phi(x_i) = \boldsymbol{C_i}\boldsymbol{\beta}_i \quad (8)$$

where $\|\boldsymbol{\beta}_i\|_0$ denotes the ℓ_0-norm that counts the nonzero components in the coefficient vector. Because of the presence of noise and possible modeling error, the optimization in (7) is often replaced by

$$\min_{\boldsymbol{\beta_i}} \|\boldsymbol{\beta_i}\|_0 \text{ subject to } \| \phi(x_i) - \boldsymbol{C_i}\boldsymbol{\beta}_i\|_2 \le \epsilon \quad (9)$$

where ϵ is an error tolerance in H. It should be noted that the nonlinear mapping function (kernel function) ϕ is unknown. The optimization problem in (9) cannot be solved directly. However, it can be replaced by

$$\min_{\boldsymbol{\beta_i}} \|\boldsymbol{\beta_i}\|_0 \text{ subject to } \| \boldsymbol{C_i}^{\boldsymbol{T}}\phi(x_i) - \boldsymbol{C_i}^{\boldsymbol{T}}\boldsymbol{C_i}\boldsymbol{\beta_i}\|_2 \le \xi \quad (10)$$

Here, any inner product value of any two high-dimensional vectors can be obtained using a kernel function k. The kernel method owes its name to the use of kernel functions, and is also called the kernel trick.

The inner product of the samples in the high-dimensional feature space can be obtained as

$$\begin{aligned}\boldsymbol{C_i^T}\phi(x_i) &= [\phi(x_1)\phi(x_i), \phi(x_2)\phi(x_i), \ldots, \phi(x_n)\phi(x_i)] \\ &= \begin{bmatrix} k(x_1, x_i) \\ k(x_2, x_i) \\ \vdots \\ k(x_n, x_i) \end{bmatrix} \\ \boldsymbol{C_i^T}\boldsymbol{C_i} &= [\phi(x_1), \ldots, \phi(x_n)]^T [\phi(x_1), \ldots, \phi(x_n)] \\ &= \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{bmatrix} \quad (11)\end{aligned}$$

Once the kernel function k is given, $\boldsymbol{C_i}^{\boldsymbol{T}}\phi(x_i)$ and $\boldsymbol{C_i}^{\boldsymbol{T}}\boldsymbol{C_i}$ are obtained. The optimization problem (9) is solved using SUnSAL [36], [46]. The reconstruction residual of x_i in H can then be acquired using the sparse vector $\boldsymbol{\beta_i}$ in H

$$\begin{aligned}r_i &= \|\phi(x_i) - \boldsymbol{C_i}\boldsymbol{\beta}_i\|_2 = \langle \phi(x_i) - \boldsymbol{C_i}\boldsymbol{\beta}_i, \phi(x_i) - \boldsymbol{C_i}\boldsymbol{\beta}_i \rangle^{\frac{1}{2}} \\ &= \left(k(x_i,x_i) - 2\boldsymbol{C}^{\boldsymbol{T}}\phi(x_i)\boldsymbol{\beta}_i + \boldsymbol{\beta_i^T}\boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{C}\boldsymbol{\beta_i}\right)^{1/2}. \quad (12)\end{aligned}$$

Just as in the original feature space for the OCC-SR, the residual vector $\boldsymbol{r} = [r_1, \ldots, r_n]$ in H is obtained when all the samples in the training dictionary have been reconstructed in H. The reconstruction residual threshold T_r is then determined by adding acceptance parameter λ. Finally, an unknown sample y is classified as belonging to the target class if it's residual in H is smaller than the given threshold.

IV. Experimental Results

We evaluated the performance of the proposed OCC method using three multispectral/hyperspectral data sets. The proposed OCC method was quantitatively compared with state-of-the-art competitors, namely, the SVDD, Parzen density estimation, *k*-means [17], and PUL classifiers [14], [25]. All these methods have been widely studied and have shown very promising results. SVDD, Parzen density estimation, and *k*-means classifiers [17] are implemented in the *dd_tools* [48] software packages.

In this paper, we used SVM classifier to estimate the probability that a positive training sample (target class sample) has been labeled (i.e., conditional probability), which is required by the PUL classifier [14], [25]. The number of unlabeled samples used is two times as the number of labeled target samples, which is similar to that adopted in [14]. Given that the Gaussian

radial basis function (RBF) has been shown to be able to handle complex nonlinear class distributions [39] and has been widely used in many studies [41], the RBF kernel was adopted in this study. In all experiments, the RBF kernel widths were tuned in the range $\sigma = \{10^{-3}, \ldots, 10^{1}\}$ for SVDD and PUL. Further, the rejection fraction parameter for SVDD, *k*-means, and Parzen density estimation methods in *dd_tools* was tuned in the range $\{10^{-3}, \ldots, 10^{0}\}$. The highest accuracy obtained was chosen as the final result for analysis. All the algorithms were implemented using MATLAB R2011 on a desktop PC equipped with an Intel Core 2 i7-2600 (at 3.4 GHz) and 8 GB of RAM memory.

In the experiments, all data were first rescaled to [0,1] and were then processed by the proposed method and other competitors. Training samples were randomly generated from the available reference data and the remaining samples were used for validation purpose. It should be noticed that, in multiclass classification, overall accuracy (OA) is usually used for the parameter optimization process. However, as we only have two classes (i.e., target and outlier classes) in OCC, the OA is meaningless especially when numbers of target class and outlier class are imbalanced. Thus, kappa coefficient was used for parameter optimization process in this paper. Moreover, the producer's accuracy (PA), user's accuracy (UA), and OA were also used to evaluate the performance of each classifier. All values of the PA, UA, OA, and kappa coefficient were obtained by averaging the values obtained after 10 Monte Carlo runs. Moreover, the level of sparsity, computed as 1-(number of atoms/number of training samples) is also given for the sparse representation-based algorithm.

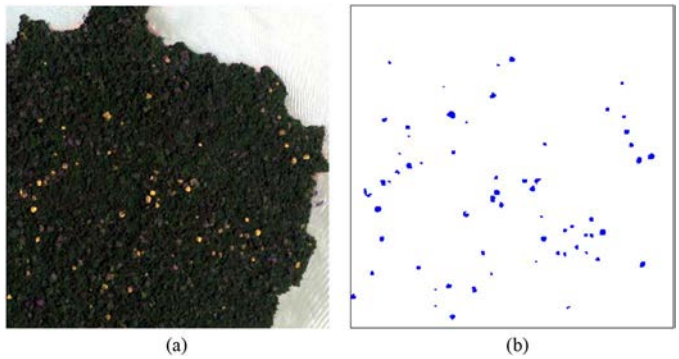

Fig. 1. QuickBird image of study area using bands 3, 4, 2 as R, G, B (a) and corresponding ground truth image (b).

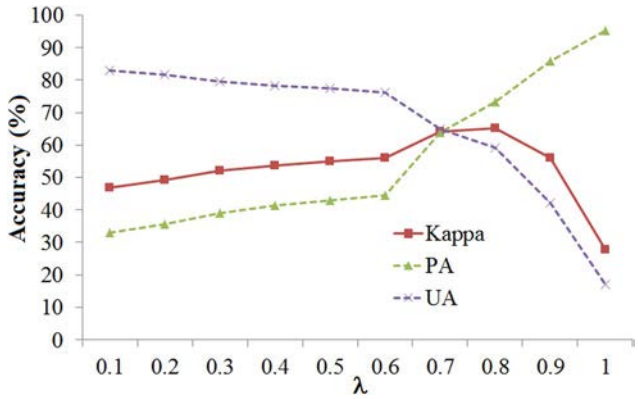

Fig. 2. Plot of kappa coefficient versus the acceptance parameter λ in OCC-SR for the QuickBird image. PA denotes the producer's accuracy and UA denotes the user's accuracy.

A. Extraction of a Specific Tree Species With QuickBird Imagery

In the first experiment, the proposed OCC methods were used to extract a specific tree species, *Tabebuia guayacan*, on the Barro Colorado Island of Panama using a QuickBird multispectral image. The *T. guayacan* tree is a hardwood tree, used extensively for construction projects since colonial times in Central America. *T. guayacan* has one of the most extensive flowering responses to precipitation after the dry season in the tropics [7]. *T. guayacan* flowering phases are "big bang" events that are triggered at the end of the dry season (February–April). Flowers grow in dense clusters and can range from 1 to 4 in in diameter. A *T. guayacan* canopy grouping presents almost 10,000 flowers in one single flowering event [7]. This event can help us to understand local ecological changes.

The QuickBird multispectral image, which consists of four multispectral bands with 2.4-m resolution, was acquired on April 29, 2002. The acquisition date of QuickBird image was selected to capture extensive flowering events of *T. guayacan* [7]. An image subset with a size of 450 $\times$ 450 pixels covering the Barro Colorado Island of Panama was used [Fig. 1(a)]. A reference map [Fig. 1(b)] is available for the image, which comprises 1492 samples representing the target class and 158,508 samples from the outlier class.

We first evaluated how the acceptance rate λ in (6) affects the classification accuracy of OCC-SR. In this experiment, 50 samples of the target class from the reference data were randomly selected as training samples and the remaining were used as test samples. The relationship between the accuracy and λ for the QuickBird image is shown in Fig. 2. From the figure, it is clear that as λ increases, PA increases and UA decreases. This is because the residual threshold T_r is very small when the λ value is small, which leads to a reconstruction residual that is larger than the given threshold for most samples. Therefore, many target class samples are misidentified. As the value of λ increases, T_r increases and more target samples are correctly classified because of larger residual thresholds, but at the same time, more outlier samples are wrongly classified as the target class, which leads to a decrease in UA. It is also clear from the figure that the kappa coefficient increases at first, and then decreases after reaching the maximum value. This is because the UA and PA are imbalanced when λ is close to 0 or 1. In other words, T_r is not optimal. Furthermore, the highest accuracy was obtained when the PA and UA were similar (balanced).

We next demonstrate the effect of λ and RBF kernel parameter γ on the performance of the proposed OCC-KSR. In each test, different λ and γ in the OCC-KSR algorithm were used to solve an OCC problem. The kappa plot for the entire test set is shown in Fig. 3. From the figure, we observe that the value of kappa increases first, and then decreases as λ increases for a given γ, which is the same behavior observed in Fig. 2. The kappa coefficient reaches a maximum value when γ equals 0.8

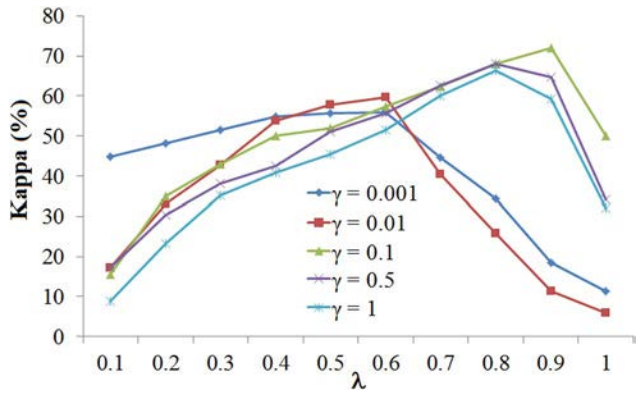

Fig. 3. Plot of kappa coefficient versus the acceptance parameter λ and γ in the kernel function of OCC-KSR for the considered QuickBird image.

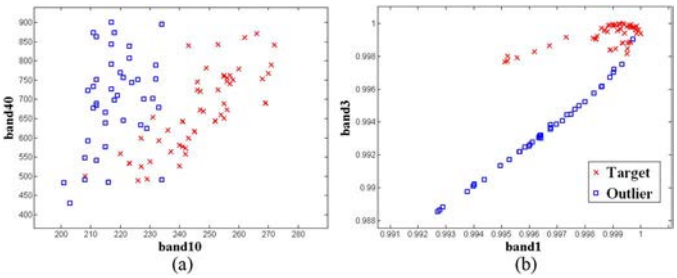

Fig. 4. Samples in the (a) original two-dimensional spectral feature space and (b) kernel feature space from Panama multispectral image.

or 0.9 in most cases. As it is also the case in other kernel-based methods [10], γ is also critical to the performance of OCC-KSR. Overall, the best classification accuracies were obtained when γ was 0.1 in this experiment.

To understand why the OCC results using kernel sparse representation (i.e., OCC-KSR) are better than those obtained by sparse representation in the original spectral space (i.e., OCC-SR), a 2-D projection of an example is shown in Fig. 4. Specifically, two scatterplots of samples from the multispectral image were constructed by randomly choosing two spectral bands in the original data [Fig. 4(a)] and in the mapping high-dimensional feature space using a kernel function [Fig. 4(b). As is clear by comparing Fig. 4(a) and (b), the class separability is improved in the mapped high-dimensional feature space, which leads to more powerful discriminative ability for OCC-KSR. Thus, OCC-KSR may achieve a better performance than OCC-SR.

A comparison between the classification results obtained by the proposed OCC method and several existing methods for the Panama QuickBird image for 10 Monte Carlo runs is given in Table I. As the results show, the proposed sparse representation OCC methods (i.e., OCC-SR and OCC-KSR) and PUL algorithm achieved higher accuracies than other comparative methods in terms of the kappa coefficient. In particular, OCC-KSR obtained the highest classification accuracy when compared with all other methods. This is probably because in high-dimensional feature space, the test samples are represented more accurately as a linear combination of training samples of the target class. For the target class (*T. guayacan* tree), SVDD, Parzen, and PUL classifier show relatively high

TABLE I
ACCURACY (THE STANDARD DEVIATION OF KAPPA IS ALSO REPORTED) OBTAINED BY DIFFERENT ONE-CLASS CLASSIFIERS AFTER EXECUTING 10 MONTE CARLO RUNS FOR THE QUICKBIRD DATA SET (ALL IN PERCENT)

Method	SVDD	k-means	Parzen	PUL	OCC-SR	OCC-KSR
PA	85.12	47.70	71.81	**90.31**	58.12	64.38
UA	39.18	**71.44**	48.27	45.98	56.12	58.22
OA	**99.23**	99.18	98.97	98.93	99.00	99.21
Kappa(std)	56.76(3.54)	56.14(8.58)	53.64(7.46)	60.41(2.21)	58.68(7.16)	**62.62**(5.26)
Sparsity					0.96	0.82

PA, producer's accuracy; UA, user's accuracy; OA, overall accuracy. Both PA and UA are for the target class.

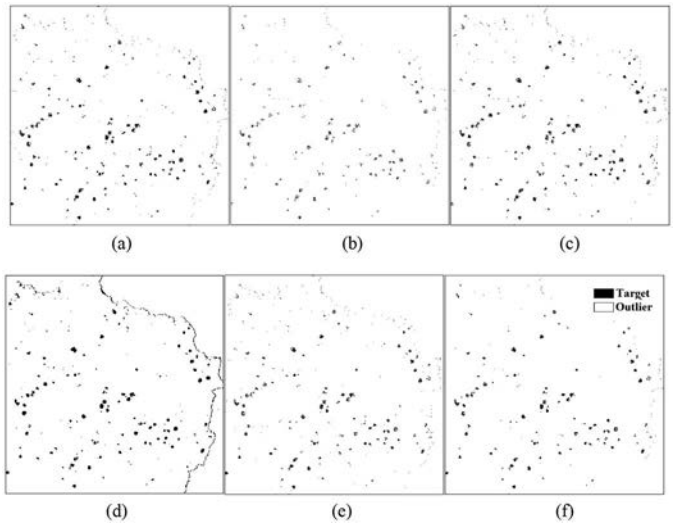

Fig. 5. Results for Panama QuickBird image. (a) SVDD. (b) k-means. (c) Parzen. (d) PUL. (e) OCC-SR. (f) OCC-KSR.

PA but relatively low UA, which implies that the target class was overestimated (i.e., high commission error). K-means classifier shows a comparatively high UA but a comparatively low PA, which implies that the target class was significantly underestimated (Table I). In contrast, our proposed methods produced more balanced and acceptable PA and high UA, which leads to a better overall classification performance.

Fig. 5 shows the classification results obtained from different methods. From the figure, it can be observed that the classification results obtained by SVDD [Fig. 5(a)] and PUL methods [Fig. 5(d)] exhibit an apparent salt-and-pepper noise in the border area between the land and sea, while some samples were wrongly classified by the *k*-means and Parzen methods [Fig. 5(b) and (c)]. However, the OCC-SR and OCC-KSR methods produced better results. In particular, the map provided by OCC-KSR is more homogeneous and exhibits a lower number of false positives. This result reveals the importance of using kernel techniques in the proposed sparse representation model.

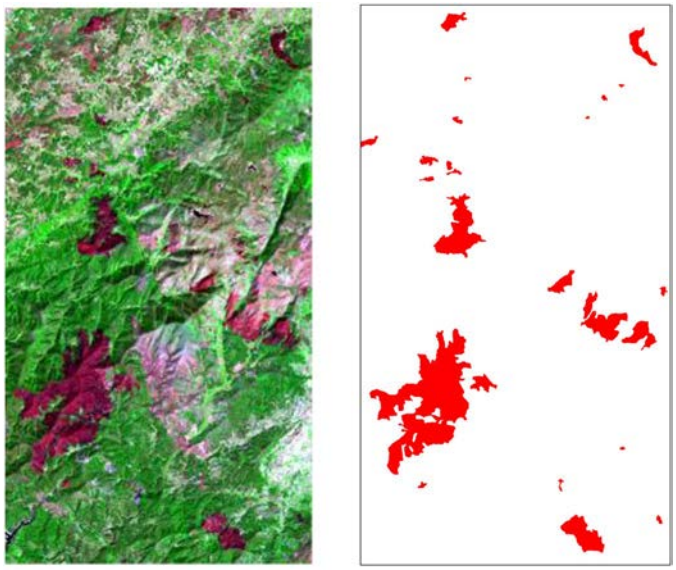

Fig. 6. False color composite image of the ETM+ image (a) and corresponding ground truth image (b).

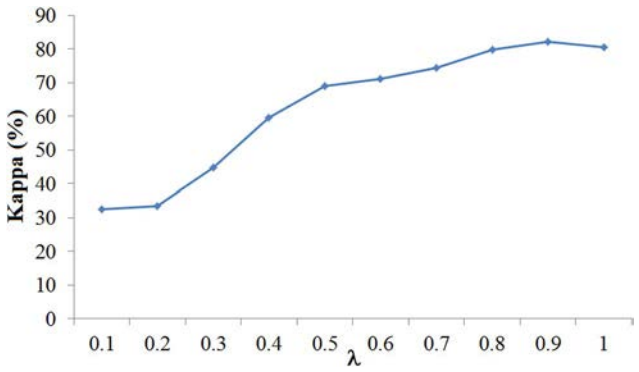

Fig. 7. Plot of kappa coefficient versus the acceptance parameter λ in OCC-SR for the Landsat ETM+ image.

B. Burned Area Mapping With Landsat ETM+ Imagery

In the second experiment, the proposed methods were evaluated for mapping burned areas caused by forest fire using a Landsat-7 ETM+ multispectral image of 30-m spatial resolution. The Landsat ETM+ image was collected over Braga, Portugal, on September 2001 [Fig. 6(a)]. The image subset is composed of 944 $\times$ 1704 pixels. The imaged area is very dry and suffers forest fires almost every summer, destroying portions of the forest and causing land cover change. Smoke and poisonous gas caused by forest fire directly or indirectly affects the environment. Thus, it is particularly important to accurately map the destroyed area to understand the local ecological change caused by forest fire.

The reference map comprises 101 953 target samples and 1 506 623 outlier samples, which were used to train the model and assess classification performance [Fig. 7(b)]. Fifty samples were randomly selected as training samples from the available reference image to build the training dictionary, and the remaining samples were used for validation.

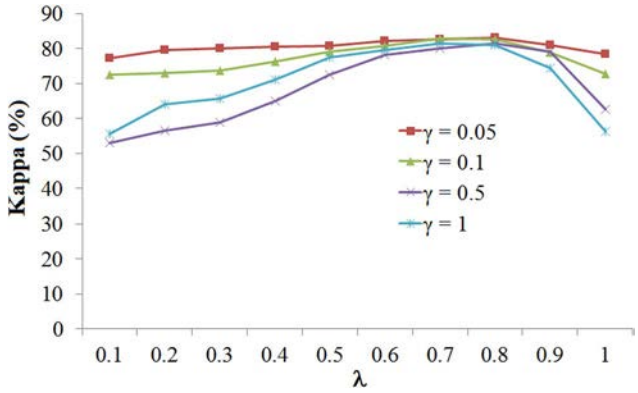

Fig. 8. Plot of kappa coefficient versus the acceptance rate λ and γ in kernel function in OCC-KSR for Landsat ETM+ image.

TABLE II
ACCURACY (THE STANDARD DEVIATION OF KAPPA IS ALSO REPORTED) OBTAINED BY DIFFERENT ONE-CLASS CLASSIFIERS AFTER EXECUTING 10 MONTE CARLO RUNS FOR THE LANDSAT ETM+DATA SET (ALL IN PERCENT)

Method	SVDD	k-means	Parzen	PUL	OCC-SR	OCC-KSR
PA	65.85	62.32	63.19	**90.87**	80.44	86.25
UA	89.75	85.21	**91.79**	71.20	86.30	82.40
OA	97.36	96.92	97.30	96.99	97.95	**97.96**
Kappa(std)	74.60(3.20)	70.41(9.92)	73.48(8.50)	78.06(5.92)	82.18(6.54)	**83.19**(5.38)
Sparsity					0.93	0.92

Both PA and UA are for the target class.

The kappa plots for the entire test set of the Landsat ETM+ image are shown in Figs. 7 and 8 for OCC-SR and OCC-KSR, respectively. It is clear that the curved shapes of the target classes in Fig. 7 display a similar trend to those shown in Fig. 3. As λ increases, the kappa coefficient of the OCC-SR first increases, then decreases after reaching a maximum value. It is clear that PA increases dramatically as λ increases, and PA and UA are balanced when λ equals 0.9, which leads to the highest accuracy obtained. From Fig. 8, it is also clear that the value of kappa first increases, then decreases with λ for a fixed γ, which is the same behavior observed in previous figures. The kappa coefficient reaches a maximum value when γ equals 0.8 in all cases. Moreover, the best classification accuracies are obtained when γ equals 0.05 in this experiment. As previously mentioned, parameters λ and γ in the Gaussian kernel function are critical to the performance of OCC-KSR.

Table II shows the classification results obtained using different OCC methods. From the table, it is clear that the proposed OCC-SR and OCC-KSR methods outperformed the considered OCC methods, with improvements of 4.12%–12.78% for kappa. For the target class (burned area), the PUL method shows a relatively high PA but relatively low UA, which implies that the target class is overestimated (i.e., high commission error). SVDD, *k*-means, and Parzen methods show relatively

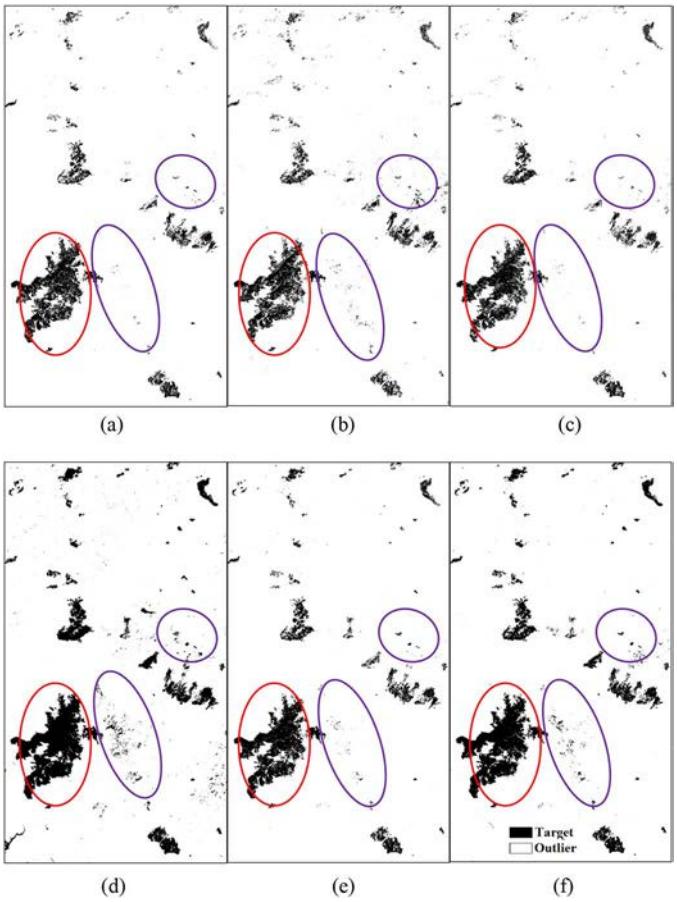

Fig. 9. Classification results for Landsat ETM+ image using (a) SVDD, (b) k-means, (c) Parzen, (d) PUL, (e) OCC-SR, and (f) OCC-KSR.

high UA but relatively low OA, which implies that the target class is significantly underestimated. In contrast, our proposed methods produced both high PA and high UA (both higher than 80%), which imply a more balanced accuracy.

The classification maps from different OCC methods in this experiment are shown in Fig. 9. In the figure, the red ellipse indicates burned area and purple ellipses indicate soil. These respective areas are shown as dark red and light red in the false color image in Fig. 6. From the classification maps [Fig. 9(e) and (f)], it is clear that the burned area (enclosed by the red ellipse) obtained from the OCC-SR and OCC-KSR approaches are homogeneous, and there are small false detections in the areas marked by purple ellipses. In comparison, there are various numbers of false detections or missed detections in the classification maps obtained by other methods. In the PUL result [Fig. 9(d)], most samples within the red ellipse were correctly classified, but clearly more false detections occurred in the purple ellipses. On the contrary, many target samples were misidentified in the red area, but less false detections occurred in the areas marked by purple ellipses in the SVDD, *k*-means, and Parzen methods [Fig. 9(a)–(c), respectively]. In conclusion, the proposed methods achieved map with good appearance in both the burned and unburned areas.

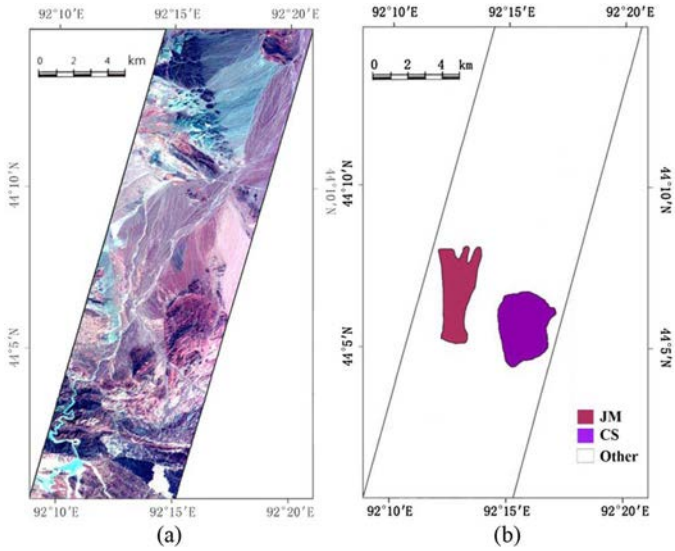

Fig. 10. False color composite image of EO-1 Hyperion image (a) and Lithological map of study area (simplified from published geologic maps at the scale of 1:200,000) (b).

C. Lithological Mapping With EO-1 Hyperion Imagery

In the third experiment, the proposed OCC methods were evaluated for lithological mapping using hyperspectral data. Lithological mapping (rock type mapping) has been an important application of hyperspectral imagery over the past three decades [49]. The hyperspectral image used in the experiment was acquired by the EO-1 Hyperion sensor on September 9, 2002. The Hyperion image comprises 242 spectral channels in the wavelength range of 0.35–2.5 μm, with a nominal spectral resolution of 10 nm and a spatial resolution of 30 m. The image covers the East Spring area, southeast of the Jung-gar Basin, Xinjiang, China (Fig. 10). The area is located in arid area with little or no vegetation, which is ideal for mapping geological information. An image subset with 545 lines by 901 samples was used (Fig. 10). The lithological units developed in the area include Carboniferous-Permian volcanic-sedimentary rocks, Jurassic sedimentary rocks, granite, and Quaternary sediments. A detailed description of these data is given in [49]. In this experiment, two lithological classes, Jurassic mudstone (JM) and Carboniferous siltstone (CS), were mapped as target classes. A reference map that comprises 1016 samples representing the JM class and 1311 samples representing the CS class was available for this experiment. As in the two previous experiments, 50 samples per target class were randomly selected from the reference data to build the training dictionary and the remaining samples (including 11 047 samples of the outlier class from the reference map) were used for testing.

Table III shows the classification results obtained from different OCC methods for the EO-1 Hyperion image. From this table, it is clear that the proposed OCC-SR produced higher or at least comparable accuracies compared with existing methods in terms of the kappa coefficient. In particular, the proposed OCC-KSR method provided the best result among all methods. We note that PA is closer to UA in the OCC-KSR results than in other methods in the experiment. Specifically, the UA is almost

TABLE III
ACCURACY (THE STANDARD DEVIATION OF KAPPA IS ALSO REPORTED) OBTAINED BY DIFFERENT ONE-CLASS CLASSIFIERS AFTER EXECUTING 10 MONTE CARLO RUNS FOR THE HYPERION DATA SET (ALL IN PERCENT)

	Class	SVDD	k-means	Parzen	PUL	OCC-SR	OCC-KSR
JM	PA	44.19	36.52	35.24	**73.84**	43.11	59.55
	UA	85.68	86.68	**90.17**	51.11	89.93	69.22
	OA	95.19	94.74	94.78	92.50	**95.30**	94.91
	Kappa(std)	56.04(4.02)	49.09(9.13)	48.47(6.28)	56.24(4.73)	56.12(5.32)	**61.29**(4.79)
	Sparsity					0.93	0.88
CS	PA	72.07	53.94	66.03	57.49	62.82	**73.37**
	UA	68.81	80.94	74.14	**81.29**	79.01	71.62
	OA	94.07	94.25	94.42	94.52	**94.73**	94.55
	Kappa(std)	67.11(3.12)	61.74(8.75)	66.78(7.35)	64.40(3.08)	67.14(6.63)	**69.46**(5.12)
	Sparsity					0.92	0.91

Both PA and UA are for the target class.

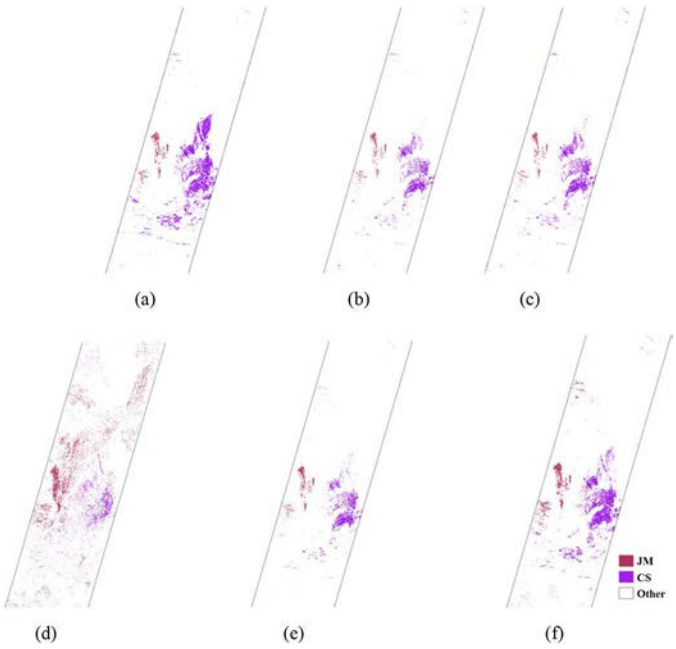

Fig. 11. Results for the considered Hyperion image. (a) SVDD. (b) k-means. (c) Parzen. (d) PUL. (e) OCC-SR. (f) OCC-KSR.

10% higher than PA in OCC-KSR for the JM class, while the difference in accuracy between PA and UA are at least 40% for other methods. This phenomenon is similar for the CS class. In other words, the OCC-KSR approach achieves a relatively higher acceptable PA and UA, which leads to a better classification performance in OCC. For illustrative purposes, Fig. 11 shows the classification maps obtained in this experiment. It is clearly found that more samples were correctly classified as JM in PUL, but at the same time, fewer outlier samples were misclassified, leading to higher PA and lower UA in comparison with other algorithms. It should be noted that the best classification accuracies were obtained when λ was 0.9 in OCC-SR, and λ was 0.7 and γ was 1 in OCC-KSR for class JM. For class CS, the best classification accuracies were obtained when λ was 0.9 in OCC-SR, and λ was 0.9 and γ was 1 in OCC-KSR.

D. General Discussion

In this paper, we proposed a new OCC method that is based on both sparse and kernel representations. From the experimental results of three study areas, the proposed OCC methods, in particular, the OCC-KSR method outperformed the considered OCC methods.

As mentioned in the previous section, an important parameter in the proposed OCC-SR and OCC-KSR methods is the acceptance rate λ, which is similar to the rejection fraction parameter in other OCC methods and used to obtain the reconstruction residual threshold. Although λ is determined by trial and error in this study, it is also found from the experimental results that the optimal threshold values of λ mainly ranged from 0.7 to 0.9 in all three study areas. This may provide a useful reference for determining the optimal value of λ. In addition, the RBF kernel width parameter γ also needs to be set in the proposed OCC-KSR. The optimal γ depends on the data set according to our experiments. However, an optimal value of γ could also be obtained by cross-validation, i.e., when the fraction of target samples that are rejected reach a given value.

The proposed OCC method has some similarities and differences with the considered OCC methods. The acceptance rate adopted in the proposed method to obtain the reconstruction residual threshold is similar to the rejection fraction parameter in other OCC methods [17]. In other words, parameter λ can be thought of as a rejection fraction parameter to help reject some unrepresentative target samples in the training. This is also very common in the other OCC approaches [17]. On the other hand, the proposed sparse representation-based OCC can also be seen as a soft version of clustering with dictionary elements being the cluster centers of the target class, i.e., each unknown (test) sample has a membership in multiple clusters, which is determined by the reconstruction residual between the sample and its approximation from element dictionary of the target class.

Our proposed sparse representation classifier provides two advantages over the considered OCC methods (SVDD, K-means, Parzen, and PUL). First, in the considered OCC methods, there is an explicit training stage and training data are only used in the training stage. A considered one-class classifier is trained only once, and then this classifier with its fixed discriminative model (or function) (e.g., support vectors for SVDD, distance for k-means, and probability for Parzen and PUL) is used to classify all of the test data. However, in the proposed OCC method, training samples are used in both the training stage to obtain a residual threshold and the classification stage to obtain a sparse representation vector for each test pixel. More importantly, a new sparse representation vector is extracted for each test pixel and is thus adaptive, representing the sparsely selected atoms that are adapted to reconstruct the current test pixel [39]. Second, to improve data separability between the target class and the outlier class, the training samples from the

target class are mapped into a high-dimensional feature space using a kernel function to build the learning dictionary. The OCC is then conducted in the new feature space. Therefore, the combination of these two advantages in the proposed method generally leads to a better performance in comparison with other existing OCC methods.

V. Conclusion

In this paper, we proposed a new OCC method for remotely sensed images using a kernel sparse representation model: OCC-KSR. The proposed approach combines the advantages of sparse representation classifiers and kernel functions to appropriately represent target samples and provide good classification results. The main contribution of our work is the integration of the sparse representation and kernel-based methods into OCC techniques for the first time. The proposed algorithms were evaluated using three multispectral/hyperspectral data sets. A comparison with state-of-the-art OCC methods demonstrated that the proposed OCC-KSR method outperformed the considered OCC methods, particularly when a very limited number of training samples were available. Future work will focus on the development of computationally efficient implementations for the techniques newly proposed in this contribution.

References

[1] C. H. Chen, *Information Processing for Remote Sensing*. Singapore: World Scientific, 1999.

[2] C. H. Chen, *Frontiers of Remote Sensing Information Processing*. Singapore: World Scientific, 2003.

[3] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support VectorMachines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[4] G. Wilkinson, "Results and implications of a study of fifteen years of satellite image classification experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 433–440, Mar. 2005.

[5] 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, Aug. 2004.

[6] J. Ham, Y. Chen, M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[7] A. Sanchez-Azofeifa *et al.*, "Estimation of the distribution of Tabebuia guayacan (Bignoniaceae) using high-resolution remote sensing imagery," *Sensors*, vol. 11, no. 4, pp. 3831–3851, 2011.

[8] C. Sanchez-Hernandez, D. Boyd, and G. Foody, "One-class classification for mapping a specific land-cover class: SVDD classification of fenland," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 1061–1073, Apr. 2007.

[9] D. Fernandez-Prieto, "An iterative approach to partially supervised classification problems," *Int. J. Remote Sens.*, vol. 23, no. 18, pp. 3887–3892, 2002.

[10] Z. Liu, W. Shi, D. Li, and Q. Qin, "Partially supervised classification–based on weighted unlabeled samples support vector machine," in *Advanced Data Mining and Applications*. New York, NY, USA: Springer, 2005, vol. 3584, pp. 118–129.

[11] A. Mazher, P. Li, and J. Zhang, "Mapping burned areas from Landsat TM imags: A comparative study," in *Proc. Int. Conf. Comput. Vis. Remote Sens.*, 2012, pp. 285–290.

[12] P. Li, H. Xuand, and J. Guo, "Urban building damage detection from very high resolution imagery using one-class SVM and spatial features," *Int. J. Remote Sens.*, vol. 31, no. 13, pp. 3393–3409, 2010.

[13] N. Longbotham *et al.*, "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009-2010 data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 331–342, Feb. 2012.

[14] W. Li, Q. Guo, and C. Elkan, "A positive and unlabeled learning algorithm for one-class classification of remote-sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 717–725, Feb. 2011.

[15] J. Munoz-Mari, F. Bovolo, L. Gomez-Chova, L. Bruzzone, and G. Camp-Valls, "Semisupervised one-class support vector machines for classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3188–3197, Aug. 2010.

[16] H. Xu and P. Li, "Land-cover change detection using one-class support vector machine," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 3, p. 255C263, 2010.

[17] D. Tax, "One-class classification," Ph.D. dissertation, Delft Univ. Technol., Delft, Netherlands, 2001.

[18] M. Kemmler, E. Rodner, and J. Denzler, "One-class classification with gaussian processes," in *Computer Vision C ACCV*. New York, NY, USA: Springer, 2011, pp. 489–500.

[19] O. Mazhelis, "One-class classifiers: A review and analysis of suitability in the context of mobile-masquerader detection," *South Afr. Comput. J.*, vol. 36, pp. 29–48, 2006.

[20] G. Cohen, H. Sax, and A. Geissbhler, "Novelty detection using one-class parzen density estimator. An application to surveillance of nosocomial infections." in *MIE*, vol. 136. Amsterdam, The Netherlands: IOS Press, 2008, pp. 21–26.

[21] D. Tax and R. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 4, pp. 45–66, 2004.

[22] J. Munoz-Mari, L. Bruzzone, and G. Camps-Valls, "A support vector domain description approach to supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2683–2692, Aug. 2007.

[23] W. Li and Q. Guo, "A maximum entropy approach to one-class classification of remote sensing imagery," *Int. J. Remote Sens.*, vol. 31, no. 8, pp. 2227–2235, 2010.

[24] C. Desir, S. Bernard, C. Petitjean, and L. Heutte, "One class random forests," *Pattern Recognit.*, vol. 46, no. 12, pp. 3490–3506, 2013.

[25] C. Elkan and K. Noto, "Learning classifiers from only positive and unlabeled data," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2008, pp. 213–220.

[26] G. M. Foody, A. Mathur, C. Sanchez-Hernandez, and D. S. Boyd, "Training set size requirements for the classification of a specific class," *Remote Sens. Environ.*, vol. 104, no. 1, pp. 1–14, 2006.

[27] W. Li and Q. Guo, "A new accuracy assessment method for one-class remote sensing classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4621–4632, Aug. 2014.

[28] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.

[29] G. Mercier and F. Girard-Ardhuin, "Partially supervised oil-slick detection by SAR imagery using kernel expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2839–2846, Oct. 2006.

[30] L. B. F. Bovolo and G. Camps-Valls, "A support vector domain method for change detection in multitemporal images," *Pattern Recognit. Lett.*, vol. 31, no. 10, pp. 1148–1154, 2010.

[31] P. Li, H. Xu, and B. Song, "A novel method of urban road damage detection using very high resolution satellite imagery and road map," *Photogramm. Eng. Remote Sens.*, vol. 77, no. 10, pp. 1057–1066, 2011.

[32] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[33] D. I. Moody, S. P. Brumby, J. C. Rowland, and C. Gangodagamage, "Undercomplete learned dictionaries for land cover classification in multispectral imagery of Arctic landscapes using CoSA: Clustering of sparse approximations," in *SPIE Defense, Security, and Sensing*. Bellingham, WA, USA: SPIE, 2013, pp. 87430B–87430B.

[34] R. Roscher and B. Waske. "Shapelet-based sparse image representation for landcover classification of hyperspectral data," in *Proc. 8th IAPR Workshop Pattern Recog. Remote Sens. (PRRS)*, 2014, pp. 1–6.

[35] Q. Sami ul Haq, L. Tao, F. Sun, and S. Yang, "A fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, Jun. 2012.

[36] B. Song *et al.*, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug. 2014.

[37] Y. Chen, N. Nasrabadi, and T. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jul. 2011.

[38] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
[39] Y. Chen, N. Nasrabadi, and T. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.
[40] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
[41] J. Yin, Z. Liu, Z. Jin, and W. Yang, "Kernel sparse representation based classification," *Neurocomputing*, vol. 77, no. 1, pp. 120–128, 2012.
[42] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
[43] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.
[44] Y. Chen, N. Nasrabadi, and T. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 676–680, Jul. 2011.
[45] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 41, no. 1, pp. 129–159, 2001.
[46] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. Piscataway, NJ, USA: IEEE, 2010, pp. 1–4.
[47] D. Ni and H. Ma, "Hyperspectral image classification via sparse code histogram," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1843–1847, Sep. 2015.
[48] D. Tax, *DDtools, the Data Description Toolbox for Matlab, Version 1.5.3*, Jun. 2006. [Online] Available: http://prlab.tudelft.nl/david-tax/dd_tools.html
[49] X. Zhang and P. Li, "Lithological mapping from hyperspectral data by improved use of spectral angle mapper," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 31, pp. 95–109, 2014.

Benqin Song received the B.E. degree in electronic and information engineering from China University of Geosciences, Beijing, China, in 2007, and the Ph.D. degree in photogrammetry and remote sensing from Peking University, Beijing, China, in 2014.

He is an Engineer with China Academy of Electronics and Information Technology, Beijing, China. His research interests include multisensor data fusion and remote sensing image classification.

Peijun Li (M'02–SM'04) received the Ph.D. degree in geology from the Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China, in 1995.

He is currently a Professor with Peking University, Beijing, China. He has been a Co-Chair of ISPRS Working Group VII/4 "Methods for land cover classification" since 2008. He was a Visiting Researcher/Professor at several institutions, including Seoul National University, Seoul, South Korea, University of Bonn, Bonn, Germany and Columbia University, New York, NY, USA. He is a Reviewer for many international journals. His research interests include land cover classification and change detection, very high resolution image analysis, urban remote sensing analysis, and geologic remote sensing.

Jun Li received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. She was a Professor with Sun Yat-sen University, Guangzhou, China. From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotcnica e de Computadores and the Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). She has also been actively involved in the Hyperspectral Imaging Network, a Marie Curie Research Training Network involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Her research interests include remote sensing, signal processing, high-performance computing, spectral unmixing, and hyperspectral image classification and segmentation.

Dr. Li is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She has been a Reviewer of several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, *Pattern Recognition*, *Optical Engineering*, *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*.

Antonio Plaza (M'05–SM'07–F'15) was born in Caceres, Spain, in 1975.

He was a Coordinator with the Hyperspectral Imaging Network, a European project with total funding of 2.8 million Euro. He has been an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, Caceres, Spain, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp), one of the most productive research groups working on remotely sensed hyperspectral data processing worldwide. He has been the advisor of 12 Ph.D. dissertations and more than 30 M.Sc. dissertations. He has authored more than 500 publications, including 162 journal papers (more than 100 in IEEE journals), 22 book chapters, and over 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He has edited a book on *High-Performance Computing in Remote Sensing* (CRC Press/Taylor & Francis) and guest edited nine special issues on hyperspectral remote sensing for different journals. He has reviewed more than 500 manuscripts for over 50 different journals. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza is an Associate Editor for the IEEE ACCESS, and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the *IEEE Geoscience and Remote Sensing Magazine* (2013). He served as the Director of Education Activities for the IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY (GRSS) in 2011–2012, and is currently serving as the President of the Spanish Chapter of IEEE GRSS (since November 2012). He has served as a Proposal Evaluator for the European Commission, the National Science Foundation, the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal. He was the recipient of the Recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009), the Recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), a journal for which he served as an Associate Editor in 2007–2012, and the Best Column Award of the *IEEE Signal Processing Magazine* in 2015, the 2013 Best Paper Award of the *JSTARS journal*, and the recipient of the most highly cited paper (2005–2010) in the *Journal of Parallel and Distributed Computing*. He was a recipient of the Best Paper Awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology, the Best Ph.D. Dissertation Award at University of Extremadura.