

## Research Article

# Sparse Representation Based Binary Hypothesis Model for Hyperspectral Image Classification

**Yidong Tang, Shucai Huang, and Aijun Xue**

*School of Air and Missile Defense, Air Force Engineering University, Xi'an 710051, China*

Correspondence should be addressed to Yidong Tang; 18109267859@163.com

Received 9 March 2016; Revised 9 May 2016; Accepted 17 May 2016

Academic Editor: Wonjun Kim

Copyright © 2016 Yidong Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The sparse representation based classifier (SRC) and its kernel version (KSRC) have been employed for hyperspectral image (HSI) classification. However, the state-of-the-art SRC often aims at extended surface objects with linear mixture in smooth scene and assumes that the number of classes is given. Considering the small target with complex background, a sparse representation based binary hypothesis (SRBBH) model is established in this paper. In this model, a query pixel is represented in two ways, which are, respectively, by background dictionary and by union dictionary. The background dictionary is composed of samples selected from the local dual concentric window centered at the query pixel. Thus, for each pixel the classification issue becomes an adaptive multiclass classification problem, where only the number of desired classes is required. Furthermore, the kernel method is employed to improve the interclass separability. In kernel space, the coding vector is obtained by using kernel-based orthogonal matching pursuit (KOMP) algorithm. Then the query pixel can be labeled by the characteristics of the coding vectors. Instead of directly using the reconstruction residuals, the different impacts the background dictionary and union dictionary have on reconstruction are used for validation and classification. It enhances the discrimination and hence improves the performance.

## 1. Introduction

The technology for artificial target recognition in nature background is important to military science and civil auto-control. Hyperspectral remote sensor captures digital images in hundreds of narrow spectral bands, which span the visible to infrared spectrum. The high spectral resolution of the data provides an invaluable source of information regarding the physical nature of the different materials and strengthens the capability to identify structures and objects in the image scene. As a result, it makes detecting and classifying the target at the same time possible, that is, integrated detection and classification. However, such a large number of spectral channels imply the high dimensionality of the data and bring challenge to image analysis. Most of the common technologies designed for the analysis of grey level, color, or multispectral images are not applicable to hyperspectral images.

One of the most important applications of HSI is classification. Different materials usually reflect electromagnetic

energy differently at specific wavelengths. This enables discrimination of materials based on the spectral characteristics. Various techniques have been developed for HSI classification. SVM [1–3] is a powerful tool to solve supervised classification problem in remote sensing image scene and performs pretty well. Variations of SVM-based algorithms have also been proposed to improve the classification accuracy [4, 5]. In the context of supervised classification, such as SVM, an additional problem is the so-called Hughes phenomenon that occurs when the training set does not have enough samples to ensure a reliable estimation of the classifier parameters. It is hard to find the separating hyperplane between two classes with very limited reference data while the number of spectral channels is usually very large in hyperspectral image scene.

The sparse representation [6] has recently been applied to hyperspectral detection and classification [7–9] relying on the observation that the pixels belonging to same class approximately lie in same low-dimensional subspace. Thus, a query pixel can be sparsely represented by a few training samples (atoms) from dictionary, and the associated sparse

representation vector will implicitly encode the class information. While SVM is a binary classifier (multiclass SVM requires one-against-one or one-against-all strategy [10]), the SRC is a multiclass classifier, which is from a reconstruction point of view. The SRC can be regarded as generalized model and often results in good performance. However, the number of classes present in hyperspectral image scene has to be known to structure the dictionary and calculate the residuals when employing the sparsity model as [11, 12]. In addition, the sparsity model may not be suitable for background pixels due to the interaction between the target and background sparse vectors.

In this paper, a sparse representation based binary hypothesis (SRBBH) model, which can be regarded as a discriminative and semisupervised model, is proposed to strengthen the performance of sparsity model. Only the number of desired classes, which is accessible in most practical cases, is required for SRBBH model. Thus, a target pixel is approximately represented by union dictionary consisting of both corresponding target training samples and background training samples while a background pixel can be approximately represented just by background dictionary. Different from SRC, different impacts the background dictionary and union dictionary have on reconstruction, instead of residuals themselves, are used for validation and classification in SRBBH model. This scheme enhances the discriminative power of different subspaces and then improves the classification performance. However, when the data structure is complex and the problem becomes nonlinear, the SRBBH model based classifier (SRBBHC) may not be competent any more. With implicitly exploiting the higher order structure of the given data, the kernel algorithm obtains significant performance improvement. Therefore, the kernel SRBBHC (KSRBBHC) is developed to project the data into high-dimensional feature space in which the data becomes linearly separable. Taking the projected data into consideration, KSRBBHC intends to separately represent the desire class and undesired class in corresponding high-dimensional feature space and makes classification performance better.

The rest of the paper is organized as follows. Section 2 briefly reviews the conventional SRC and its kernel version. Section 3 proposes the SRBBHC and its kernel version for HSI. The effectiveness of proposed SRBBHC and KSRBBHC is demonstrated by experimental results in Section 4. Finally, conclusions are drawn in Section 5.

## 2. Sparse Representation Based Classification

**2.1. SRC.** In sparsity model, it is assumed that the spectral signatures of pixels belonging to same class approximately lie in same low-dimensional subspace. A query pixel is given  $\mathbf{y} \in \mathbb{R}^B$ , where  $B$  is the number of bands. Then the linear representation of  $\mathbf{y}$  can be written in terms of all training samples as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (1)$$

where  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_M] \in \mathbb{R}^{B \times N}$  is a structured dictionary whose columns are  $N$  training samples of all  $M$

classes and  $\mathbf{x} \in \mathbb{R}^{N \times 1}$  is the sparse coefficient vector.  $\mathbf{x}$  can be recovered by solving

$$\begin{aligned} \hat{\mathbf{x}} &= \arg \min \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \\ \text{s.t. } &\|\mathbf{x}\|_0 \leq K_0, \end{aligned} \quad (2)$$

where  $\|\cdot\|_0$  denotes the  $l_0$ -norm, which is defined as the number of nonzero entries in the vector, and  $K_0$  is a preset upper bound on sparsity level. The problem in (2) is a NP-hard problem, which can be approximately solved by greedy algorithms, such as orthogonal matching pursuit (OMP) [13] and subspace pursuit (SP) [14], or relaxed to convex programming [15]. In this paper, the OMP algorithm is exploited to generate sparse coefficient vector. The OMP algorithm augments the support set by one index per iteration until  $K_0$  atoms are selected or the approximation error is within a preset threshold.

Once the sparse coefficient vector is obtained, the class label of  $\mathbf{y}$  is determined by the minimal residual between  $\mathbf{y}$  and its approximation from each class of subdictionary:

$$\text{Class}(\mathbf{y}) = \arg \min_{i=1, \dots, M} \|\mathbf{y} - \mathbf{A}_i \hat{\mathbf{x}}_i\|_2, \quad (3)$$

where  $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \dots, \mathbf{a}_{i,n_i}] \in \mathbb{R}^{B \times n_i}$  is the dataset of training sample from class  $i$  and  $\hat{\mathbf{x}}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,n_i}]^T$  is the coefficient vector associated with  $\mathbf{A}_i$ .

**2.2. KSRC.** Kernel methods outperform the classical linear algorithms by implicitly exploiting the nonlinear information of given data [16]. It relies on the observation that a pixel in kernel-induced feature space can be linearly represented in terms of the training samples in same space [9]. Let  $\mathbf{y} \in \mathbb{R}^B$  be the data point of interest and let  $\phi(\mathbf{y})$  be its representation in kernel-induced feature space. Similar to the SRC, the linear representation of  $\phi(\mathbf{y})$  in terms of training samples in kernel-induced feature space can be formulated as

$$\begin{aligned} \phi(\mathbf{y}) &= \underbrace{[\phi(\mathbf{A}_1), \dots, \phi(\mathbf{A}_M)]}_{\mathbf{A}^\phi} \underbrace{[\phi(\mathbf{x}_1)^T, \dots, \phi(\mathbf{x}_M)^T]^T}_{\mathbf{x}^\phi} \\ &= \mathbf{A}^\phi \mathbf{x}^\phi, \end{aligned} \quad (4)$$

where  $\mathbf{A}^\phi$  is the training dictionary in kernel-induced feature space and  $\mathbf{x}^\phi$  is the coefficient vector. The vector  $\mathbf{x}^\phi$  can be recovered by solving

$$\begin{aligned} \hat{\mathbf{x}}^\phi &= \arg \min \|\mathbf{y} - \mathbf{A}^\phi \mathbf{x}^\phi\|_2 \\ \text{s.t. } &\|\mathbf{x}^\phi\|_0 \leq K_0. \end{aligned} \quad (5)$$

Problem (5) can be approximately solved by kernelized sparse recovery algorithms, such as the kernelized orthogonal matching pursuit (KOMP) and kernelized subspace pursuit (KSP). Implementation details of the KOMP and KSP can be found in [17]. In this paper, the KOMP is used to solve the problem with RBF kernel function. To avoid directly evaluating the inner product in high-dimensional feature

space, the kernel-based learning algorithm uses an effective kernel trick to implement dot products in the feature space without knowing the exact mapping function  $\phi$ .

Once the sparse vector  $\hat{\mathbf{x}}^\phi$  is obtained, the residual associated with  $i$ th subject in the feature space is then computed by

$$\begin{aligned} r_i(\mathbf{y}) &= \|\phi(\mathbf{y}) - \mathbf{A}_i^\phi \hat{\mathbf{x}}_i^\phi\| \\ &= \sqrt{\left( k(\mathbf{y}, \mathbf{y}) - 2(\hat{\mathbf{x}}_i^\phi)^T k(\mathbf{A}_i^\phi, \mathbf{y}) + (\hat{\mathbf{x}}_i^\phi)^T k(\mathbf{A}_i^\phi, \mathbf{A}_i^\phi) \hat{\mathbf{x}}_i^\phi \right)} \quad (6) \\ &= \sqrt{\left( k(\mathbf{y}, \mathbf{y}) - 2(\hat{\mathbf{x}}_i^\phi)^T (\mathbf{k}_{A,y})_i + (\hat{\mathbf{x}}_i^\phi)^T (\mathbf{K}_A)_{i,i} \hat{\mathbf{x}}_i^\phi \right)}, \end{aligned}$$

where  $\mathbf{k}_{A,y}$  and  $\mathbf{K}_A$  are, respectively, kernel tricks of target dictionary with the query pixel and itself. The class label of query pixel  $\mathbf{y}$  is then determined as

$$\text{Class}(\mathbf{y}) = \arg \min_{i=1,\dots,M} r_i(\mathbf{y}). \quad (7)$$

Though SRC and KSRC have been proved to be powerful approaches as shown in [8], the main idea of the SRC and KSRC is only appropriate for extended surface target classification such as plantation and geology. In this case, the pixels in a large neighborhood are likely to consist of similar materials, and different subjects are next to each other without undesired subject (background) existing between them. As a result, the spectral mixture only occurs along the boundary, leading to the fact that most of the target pixels are observed without corruption brought by background and most of the training samples selected from dataset for dictionary are pure. However, this probably would never happen to small size target whose spectrum is almost mixed with background. Thus, it may be unreliable for conventional SRC and KSRC to use the residual information for validation and classification. On the other hand, it is assumed that we have hold all class of subjects present in the hyperspectral image scene, including the number of classes and corresponding training samples. In other words, the SRC and KSRC cannot distinguish between small targets in hyperspectral image scene for generally lacking training samples of undesired class.

### 3. SRBBHC and the Kernelized Version

In this section, we introduce the proposed SRBBH model based classification algorithm for HSI, which utilizes the binary hypothesis for quality validation as well as the reconstruction residuals by the two hypotheses for classification. Moreover, a kernelized version of the proposed classifier is also introduced for nonlinear classification in a high-dimensional feature space.

**3.1. SRBBHC.** When employing SRC and KSRC for HSI, it is assumed that the number of classes present in the image scene  $M$  is known. However, it is actually difficult to know this information due to scene complexity. In many practical situations, only the number of desired classes is available.

Fortunately, considering the regions of interest, such as artificial target in nature background, what we wonder is the class label of desired subjects, but not the class label of all kinds of subjects. In other words, we need to detect and then reject the undesired query sample before classification. On the other hand, different from extended surface target, the pixels of target with small size are almost mixed with background spectrum. Detecting and classifying desired subjects from the mixed pixel are generally difficult, especially when the background spectrum has a close or even larger abundance than target. In addition, although the background and target training samples have distinct spectral signatures and lie in two different subspaces, the two subspaces are usually not orthogonal, due to spectral variation [18]. In such case, the reconstruction residual via corresponding target training samples may be on the contrary larger than that via background training samples, which will lead to mistake target for background. Thus, it is no longer sufficient to directly use the reconstruction residuals for validation and classification. The SRBBHC solves these problems by utilizing a binary hypothesis model with more reasonable dictionaries, where the query pixel is, respectively, modeled with background dictionary under the null hypothesis and with union dictionary under the alternative hypothesis. And then the binary hypothesis is used for validation. In SRBBHC and its kernelized version, the different impacts the background dictionary and union dictionary have on reconstruction, instead of residuals themselves, are used for validation and classification. In a sense, the SRBBHC can be viewed as a joint target detection and classification scheme, which firstly detects the valid samples and then classifies them.

In detail, denote the union dictionary consisting of both target training samples from class  $i$  and background training samples as  $\mathbf{A}_i^u = [\mathbf{A}_i^t, \mathbf{A}^b]$ ,  $i = 1, \dots, M$ , where  $\mathbf{A}_i^t$  is target subdictionary associated with class  $i$  and  $\mathbf{A}^b$  is background subdictionary. If  $\mathbf{y}$  belongs to undesired class, the spectrum lies in a low-dimensional subspace spanned by the background training samples. As a result, the residuals of different union dictionary  $\mathbf{A}_i^u$  are similar to the residual of background subdictionary  $\mathbf{A}^b$ , that is, hypothesis  $H_0$ . On the other hand, if  $\mathbf{y}$  belongs to class  $i$ , the union dictionary  $\mathbf{A}_i^u$  will give better representation, leading to smaller residual than background subdictionary, that is, hypothesis  $H_1$ . The binary hypothesis for quality validation is modeled as follows:

$$\begin{aligned} H_0 &: \|\mathbf{y} - \mathbf{A}^b \boldsymbol{\gamma}\| \\ &\approx \|\mathbf{y} - \mathbf{A}_i^u \boldsymbol{\beta}_i\|_{i=1,2,\dots,M}, \text{ undesired class} \quad (8) \\ H_1 &: \|\mathbf{y} - \mathbf{A}^b \boldsymbol{\gamma}\| > \|\mathbf{y} - \mathbf{A}_i^u \boldsymbol{\beta}_i\|_{i=1,2,\dots,M}, \text{ desired class,} \end{aligned}$$

where  $\boldsymbol{\gamma}$  and  $\boldsymbol{\beta}_i$  are coefficient vectors associated with  $\mathbf{A}^b$  and  $\mathbf{A}_i^u$ , respectively. In other words, the problem is reformed into local binary classification problem, where the binary hypothesis is used to decide if the test pixel is a valid sample from one of the classes we desire.

According to the sparse coding theory [15], the coefficient vectors can be recovered by solving following minimization problem with the same sparsity level:

$$\begin{aligned} \hat{\boldsymbol{\gamma}} &= \arg \min \quad \|\mathbf{x} - \mathbf{A}^b \boldsymbol{\gamma}\|_2 \\ \text{s.t.} \quad & \|\boldsymbol{\gamma}\|_0 \leq K_0, \\ \hat{\boldsymbol{\beta}}_m &= \arg \min \quad \|\mathbf{x} - \mathbf{A}_i^u \boldsymbol{\beta}_i\|_2 \\ \text{s.t.} \quad & \|\boldsymbol{\beta}_i\|_0 \leq K_0. \end{aligned} \quad (9)$$

Once the sparse coefficient vectors are obtained, the semantic information can be directly extracted from the coefficient vectors. The residuals of background subdictionary and different union dictionaries are calculated as

$$\begin{aligned} r_0 &= \|\mathbf{y} - \mathbf{A}^b \hat{\boldsymbol{\gamma}}\|, \\ r_i &= \|\mathbf{y} - \mathbf{A}_i^u \hat{\boldsymbol{\beta}}_i\|_{i=1,2,\dots,M}. \end{aligned} \quad (10)$$

If we decide the given  $\mathbf{y}$  as a valid sample belongs to class  $i$ , the union dictionary  $\mathbf{A}_i^u$  will also give much better representation than the other union dictionaries  $\mathbf{A}_{j,j \neq i}^u$ , leading to larger difference between  $r_i$  and  $r_0$ . Then,  $\mathbf{y}$  will be labeled to the class with greatest difference between  $r_i$  and  $r_0$ . Defining a vector  $\mathbf{R} = (r_0 - r_1, r_0 - r_2, \dots, r_0 - r_M)$ , the outputs of integrated detection and classification decision are then made by

$$\begin{aligned} \text{Detector}(\mathbf{y}) &= \max \quad (\mathbf{R}) \\ \text{Class}(\mathbf{y}) &= \arg \max_{i=1,\dots,M} (r_0 - r_i) \\ \text{s.t.} \quad & \text{Detector}(\mathbf{y}) \geq \delta, \end{aligned} \quad (11)$$

where  $\delta$  is a threshold used for validation. When  $\text{Detector}(\mathbf{y}) < \delta$ , the query pixel will be labeled as undesired class, that is, background. The threshold  $\delta$  makes important effects on validation and hence classification. However, in this study, the threshold  $\delta$  is determined experimentally due to lack of parameter analysis theory. In our future work, we will investigate how to automatically choose appropriate  $\delta$  for different test datasets.

Considering the size of the desired subjects, the background dictionary is generated locally for each query pixel through a dual concentric window centered at query pixel. Only the samples in the outer region are involved in  $\mathbf{A}^b$ . As a result, the background dictionary is constructed adaptively for each pixel and captures the background spectral signature of the query pixel better. It is important to note that same sparsity level must be adopted for each union dictionary  $\mathbf{A}_i^u$  to make sure of the comparability of residuals of two hypotheses in SRBBHC.

**3.2. KSRBBHC.** For a hyperspectral image scene, the spectral mixing may be nonlinear due to the complex imaging condition in many practical situations [19]. As a result, the data structure may become complex and the problem becomes no longer linearly separable. In such case, the linear SRBBHC is not competent any more. Fortunately, kernel methods can project the linearly nonseparable data into a high-dimensional feature space in which those data become more separable. Here we extend the proposed SRBBHC into a kernel vision, referred to as KSRBBHC.

Similar to the KSRC, suppose that  $\phi(\mathbf{y})$  is the representation of query pixel in the high-dimensional feature space; the SRBBH model becomes

$$\begin{aligned} H_0 &: \|\phi(\mathbf{y}) - \mathbf{A}^{b\phi} \boldsymbol{\gamma}^\phi\| \\ &\approx \|\phi(\mathbf{y}) - \mathbf{A}_i^{u\phi} \boldsymbol{\beta}_i^\phi\|_{i=1,2,\dots,M}, \text{ undesired class} \\ H_1 &: \|\phi(\mathbf{y}) - \mathbf{A}^{b\phi} \boldsymbol{\gamma}^\phi\| \\ &> \|\phi(\mathbf{y}) - \mathbf{A}_i^{u\phi} \boldsymbol{\beta}_i^\phi\|_{i=1,2,\dots,M}, \text{ desired class,} \end{aligned} \quad (12)$$

where  $\mathbf{A}^{b\phi}$  and  $\mathbf{A}_i^{u\phi}$ , respectively, result from mapping  $\mathbf{A}^b$  and  $\mathbf{A}_i^u$  into kernel-induced feature space by mapping function  $\phi$  and  $\boldsymbol{\gamma}^\phi$  and  $\boldsymbol{\beta}_i^\phi$  are coefficient vectors, respectively, associated with  $\mathbf{A}^{b\phi}$  and  $\mathbf{A}_i^{u\phi}$ .

Employing the RBF kernel function, the residuals of the kernel SRBBHC in terms of background dictionary and different union dictionaries are, respectively, computed by

$$\begin{aligned} r_0^\phi &= \|\phi(\mathbf{y}) - \mathbf{A}^{b\phi} \hat{\boldsymbol{\gamma}}^\phi\| = \sqrt{\left( k(\mathbf{y}, \mathbf{y}) - 2 (\hat{\boldsymbol{\gamma}}^\phi)^T k(\mathbf{A}^{b\phi}, \mathbf{y}) + (\hat{\boldsymbol{\gamma}}^\phi)^T k(\mathbf{A}^{b\phi}, \mathbf{A}^{b\phi}) \hat{\boldsymbol{\gamma}}^\phi \right)}, \\ r_i^\phi &= \|\phi(\mathbf{y}) - \mathbf{A}_i^{u\phi} \hat{\boldsymbol{\beta}}_i^\phi\| = \sqrt{\left( k(\mathbf{y}, \mathbf{y}) - 2 (\hat{\boldsymbol{\beta}}_i^\phi)^T k(\mathbf{A}_i^{u\phi}, \mathbf{y}) + (\hat{\boldsymbol{\beta}}_i^\phi)^T k(\mathbf{A}_i^{u\phi}, \mathbf{A}_i^{u\phi}) \hat{\boldsymbol{\beta}}_i^\phi \right)}, \end{aligned} \quad (13)$$

where  $\hat{\boldsymbol{\gamma}}^\phi$  and  $\hat{\boldsymbol{\beta}}_i^\phi$  are the estimation for  $\boldsymbol{\gamma}^\phi$  and  $\boldsymbol{\beta}_i^\phi$  by KOMP.

Similarly,  $\mathbf{y}$  will be labeled to the class with greatest difference between  $r_i^\phi$  and  $r_0^\phi$ . Defining a new vector  $\mathbf{R}^\phi = (r_0^\phi - r_1^\phi, r_0^\phi - r_2^\phi, \dots, r_0^\phi - r_M^\phi)$ , the outputs of integrated detection and classification decision are then made by

$$\begin{aligned} \text{Detector}(\mathbf{y}) &= \max \quad (\mathbf{R}^\phi) \\ \text{Class}(\mathbf{y}) &= \arg \max_{i=1,\dots,M} (r_0^\phi - r_i^\phi) \\ \text{s.t.} \quad & \text{Detector}(\mathbf{y}) \geq \delta^\phi, \end{aligned} \quad (14)$$

where  $\delta^\phi$  is also a threshold used for validation. When  $\text{Detector}(\mathbf{y}) < \delta^\phi$ , the query pixel will be labeled as undesired class, that is, background. Similar to SRBBHC, although  $\delta^\phi$  is important for validation and classification, it is also determined experimentally.

#### 4. Experimental Results

In this section, the classification performance of KSRBBHC is evaluated and compared to the other four classifiers (SVMC, SRC, KSRC, and SRBBHC), and RBF kernel function  $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma^2)$  is used for KSRC and KSRBBHC. The average recognition rate (ARR) and overall recognition rate (ORR) are suggested as performance parameters. The effectiveness of the proposed algorithms is evaluated with two datasets: a synthetic dataset ROI-I and a real dataset ROI-II, as shown in Figures 4 and 5.

The ROI-I is constructed by implanting five classes of targets, which are, respectively, *artificiality*, *clay*, *tree*, *plane*, and *grass* with a background scene size of  $100 \times 100$  pixels collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) from San Diego, CA, USA. The image has 224 spectral channels (189 available) in wavelengths ranging from 370 to 2510 nm. In detail, each class of target is linearly mixed with background by varying abundance from 0.3 to 0.7 with step 0.1 in five small neighborhoods of size  $2 \times 2$  pixels. In other words, the image contains 25 desired subjects occupying 100 pixels. 4 unmixed samples per class are randomly chosen for training, and the dual window sizes  $\omega_{\text{in}}$ ,  $\omega_{\text{out}}$  are set as  $3 \times 3$  and  $9 \times 9$  according to the size of desired target. We compute the ARR and ORR of KSRC and KSRBBHC for ROI-I with varying kernel parameter  $\sigma$  and sparsity level  $K_0$ , as shown in Figure 1. One can see from Figure 1 that the KSRC and KSRBBHC both are sensitive to  $\sigma$ , while  $K_0$  plays a nearly negligible role for the two kernel vision algorithms when  $\sigma$  is fixed, especially when  $\sigma$  is fixed as 0.001, 0.01, and 0.1. When  $\sigma$  is fixed as 10, the ARR and ORR of KSRC and KSRBBHC both remain at a relatively high level and change smoothly with  $K_0$ . As a result, the kernel parameter  $\sigma$  is set as 10 for the KSRC and KSRBBHC.

The ROI-II with the size of  $100 \times 100$  pixels is a region directly taken from the real AVIRIS image, San Diego. It contains 3 classes of desired subjects occupying 317 pixels and undesired subjects occupying 9683 pixels. For each class, around 10% of the labeled samples are chosen randomly for training. The dual window sizes  $\omega_{\text{in}}$ ,  $\omega_{\text{out}}$  are set as  $7 \times 7$  and  $11 \times 11$  manually according to our previous work [20]. The ARR and ORR of KSRC and KSRBBHC with varying  $\sigma$  and  $K_0$  for ROI-II are shown in Figure 2. In detail, as shown in Figure 2(a), when  $\sigma$  is very small (0.001, 0.01) the KSRC always has a poor performance. When  $\sigma$  is relatively large, in general, the performance of KSRC increases with  $K_0$  smoothly, until the parameter reaches a certain level. Only when  $\sigma$  is too large ( $\sigma = 100$ ), the performance of KSRC fluctuates with the value of  $K_0$ . For KSRBBHC, as shown in Figure 2(b), the performance remains at high level at all sparsity level when  $\sigma$  is small (0.001, 0.01, and 0.1). And the performance increases with  $K_0$  smoothly when  $\sigma$  is fixed as

1, until the parameter reaches a certain level. However, the performance decreases with  $K_0$  when  $\sigma$  is large (10, 100). In a word, the experiment results for ROI-II also show that the classification performance is more sensitive to kernel parameter rather than sparsity level. Unlike the same optimal kernel parameter setting in ROI-I, the kernel parameter  $\sigma$  is, respectively, set as 10 for the KSRC and 0.01 for KSRBBHC in ROI-II.

Figure 3 shows the classification performances of the four sparsity-based classifiers with varying  $K_0$  when kernel parameter  $\sigma$  is, respectively, optimized for corresponding algorithm. For the ROI-I, as shown in Figure 3(a), in general, the experiment result shows that the classification performances of four algorithms increase with  $K_0$ , until the parameter reaches a certain level, and then the performances hold the line. As a result, the parameter  $K_0$  is set as 8 to balance the performance and complexity for all sparsity-based classifiers in ROI-I. For the ROI-II, the classification performances of SRC, KSRC, and SRBBHC increase with  $K_0$  when the parameter is smaller than 6; then the performances of SRC and KSRC keep increasing smoothly with  $K_0$ , whereas the performance of SRBBHC reduces gradually. The reason for this may be that only proper  $K_0$  can lead a good discriminative performance of sparsity-based classifiers. In detail, for a very small  $K_0$ , the residuals of background dictionary and union dictionary are both large, which finally weakens the classification performance. When  $K_0$  is too large, along with the nonorthogonality between desired target and background, the solution may be dense and result in a degraded discriminative power. Exactly as expected, for both ROI-I and ROI-II the performance of KSRBBHC reaches a good result with relatively small  $K_0$  and remains at high level when  $K_0$  increases (i.e., insensitive to  $K_0$ ). The parameter  $K_0$  is set as 6 to balance the performance and complexity for all sparsity-based classifiers in ROI-II.

After getting the optimal values of  $K_0$  and  $\sigma$ , the experimental results of the five classifiers, averaged over five independent realizations, are determined via ORR and ARR, as shown in Tables 1 and 2. Because only the number of desired classes is known, it should be noted that the SVMC is designed as a local  $M + 1$ -class ( $M = 5$  for ROI-I;  $M = 3$  for ROI-II) classifier with one-against-all strategy, where the training samples from class  $M + 1$  (background) are adaptively collected for each query pixel in local dual concentric window. One can observe from Tables 1 and 2 that both SRBBHC and KSRBBHC have significantly improved the classification accuracy. One can also see that the relationship of ORR and that of ARR between two classifiers are sometimes mismatched. The reason for this phenomenon may be that the total number of desired pixels is so small that even modest increase of desired target classification accuracy can lead to enormous increase on ARR but only small change on ORR. For example, for the KSRBBHC in Table 1, the classification accuracy for every single class is greater than or equal to that of SRBBHC, leading to a greater ARR. However, due to the small ratio but big absolute amount growth of classification on background, the ORR of SRBBHC is turning to greater than KSRBBHC. Considering the mismatch, the best one of the five independent results for each classifier

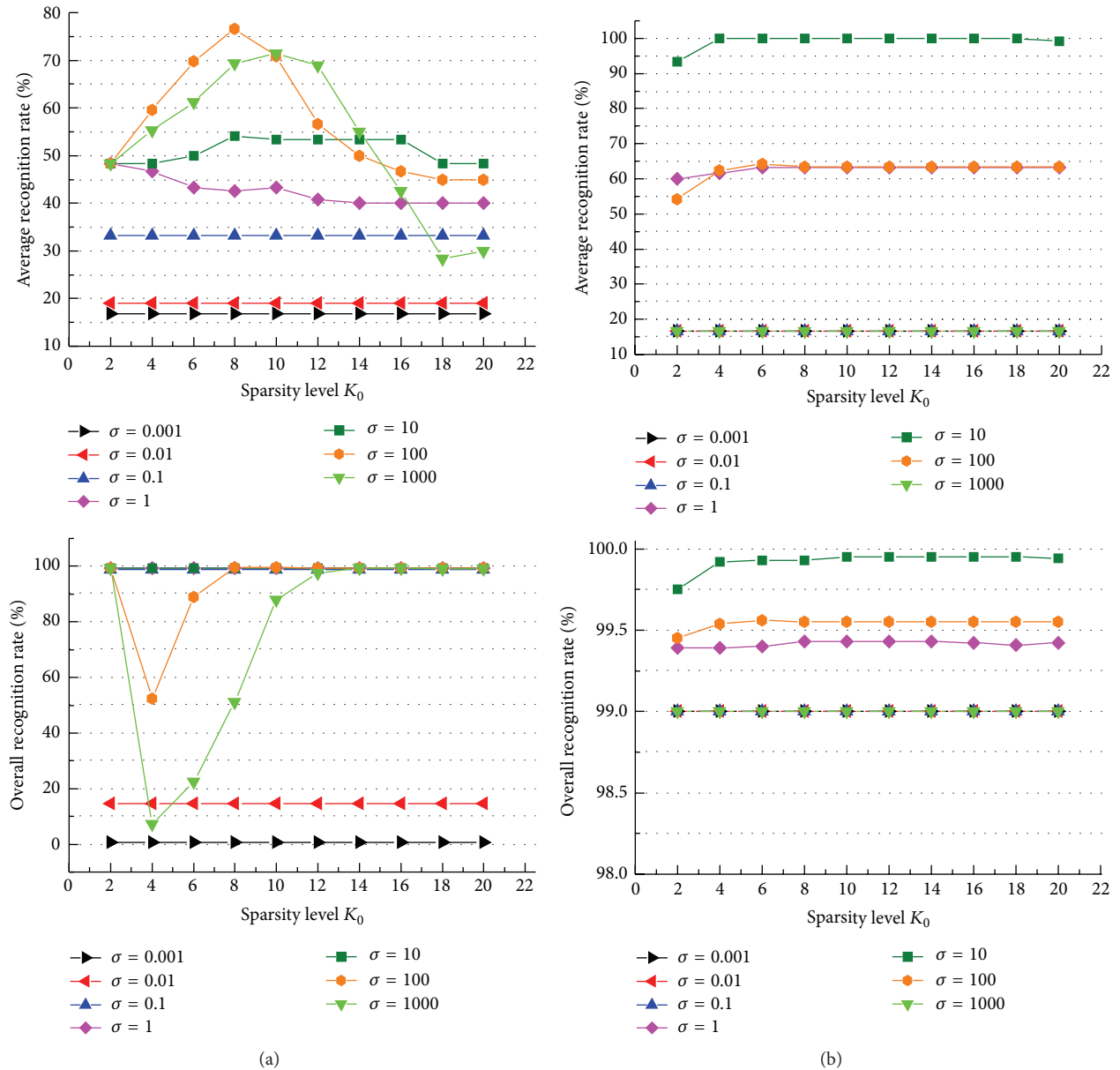


FIGURE 1: Effect of sparsity level  $K_0$  and RBF kernel parameter  $\sigma$  on performance of two kernel-based classifiers: (a) KSRC; (b) KSRBBHC for ROI-I.

is chosen to estimate the performance intuitively, as shown in Figures 4 and 5. The corresponding algorithm and ORR are presented on the bottom of each map. One can clearly see that both binary hypothesis model based classifiers have significantly improved classification performance. The SRBBHC not only classified the desired target exactly but also avoid mistaking background to desired target effectively. Furthermore, the KSRBBHC outperforms all other classifiers. In detail, for ROI-I, because of the mixture of desired target and background, the SVMC, SRC, and KSRC mistake desired target to background to a great extent. Severely, the SRC mistakes abundant background pixels to desired target, while the KSRBBHC and SRBBHC yield a superior performance. For ROI-II, the SVMC barely classify the 2nd class of target

correctly because of lack of enough training sample and the similarity between 2nd class of target and background. And the conventional sparsity-based classifiers (SRC and KSRC) lead to quite a number of misclassifications due to their weak discriminative power. In a word, one can observe from the experimental results that the SRBBH model based classifier (SRBBHC) offers better performance than some traditional techniques. Moreover, the kernelization of SRBBHC further improves the performance.

## 5. Conclusion

In this paper, a sparse representation based binary hypothesis model is proposed for target classification in HSI. And the

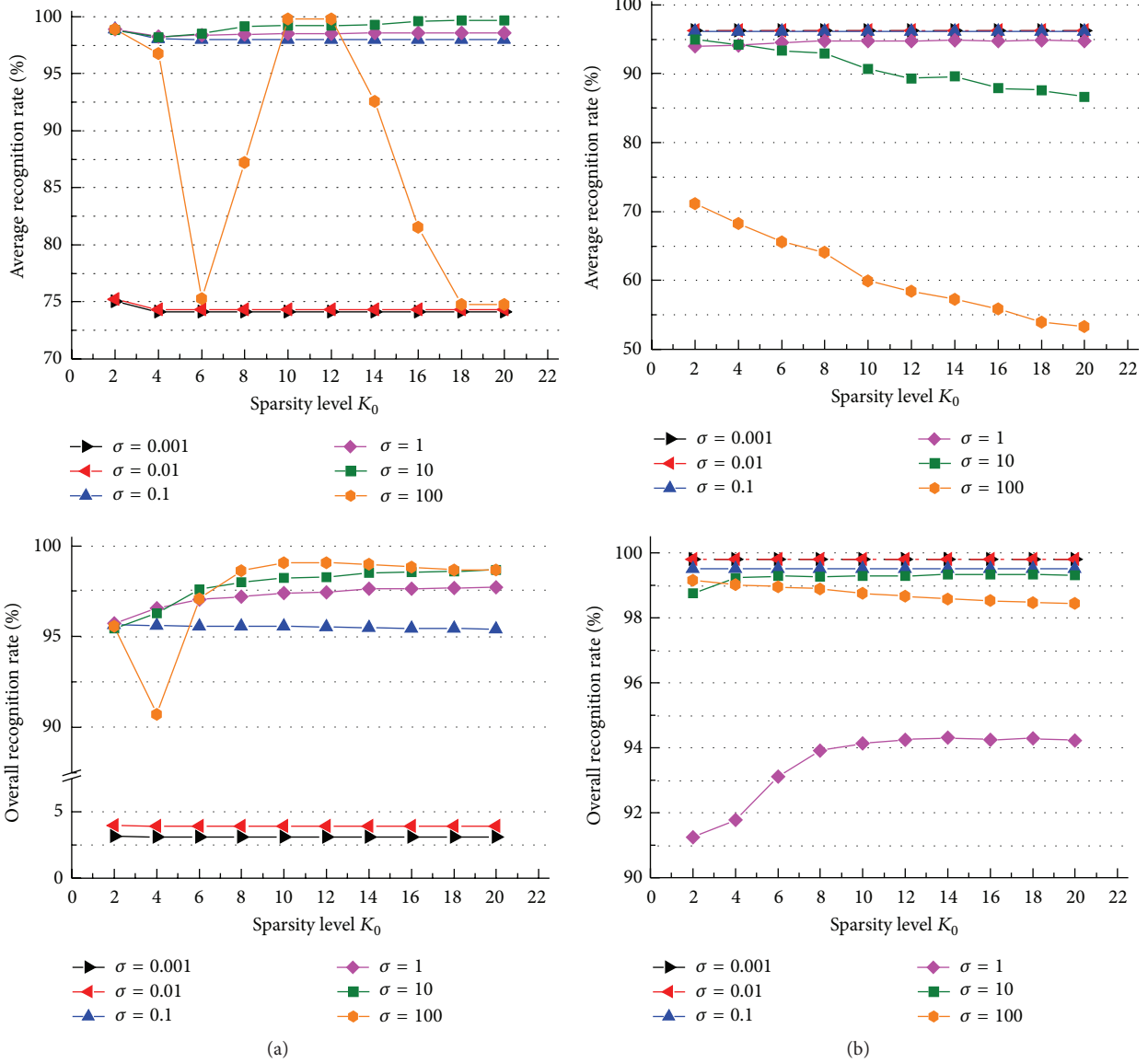


FIGURE 2: Effect of sparsity level  $K_0$  and kernel parameter  $\sigma$  on classification performance of two kernel classifiers: (a) KSRC; (b) KSRBBHC for ROI-II.

TABLE 1: Classification accuracy for the ROI-I using 4 unmixed samples as training set.

Class	SVMC	SRC	KSRC	SRBBHC	KSRBBHC
1	19	73	59	<b>100</b>	<b>100</b>
2	0	55	36	<b>100</b>	<b>100</b>
3	64	3	36	<b>100</b>	<b>100</b>
4	35	54	48	<b>100</b>	<b>100</b>
5	80	56	61	<b>100</b>	<b>100</b>
Background	100	91.42	<b>100</b>	99.73	99.92
ORR	99.40	86.54	99.49	99.73	<b>99.92</b>
ARR	49.67	59.01	57.33	99.95	<b>99.98</b>
Time	172.96	<b>11.35</b>	15.28	33.03	72.71

TABLE 2: Classification accuracy for the ROI-II using around 10% labeled samples as training set.

Class	SVMC	SRC	KSRC	SRBBHC	KSRBBHC
1	99.64	99.82	<b>100</b>	89.01	91.62
2	5.38	<b>100</b>	<b>100</b>	92.31	94.13
3	<b>100</b>	98.65	98.60	93.24	93.22
Background	98.52	98.16	98.15	99.81	99.69
ORR	97.66	96.81	97.55	98.65	<b>99.07</b>
ARR	75.38	97.20	<b>98.05</b>	91.46	94.11
Time	471.45	23.91	20.15	<b>14.69</b>	30.91

corresponding algorithm (SRBBHC) based on this model is proposed for HSI classification. Furthermore, taking the

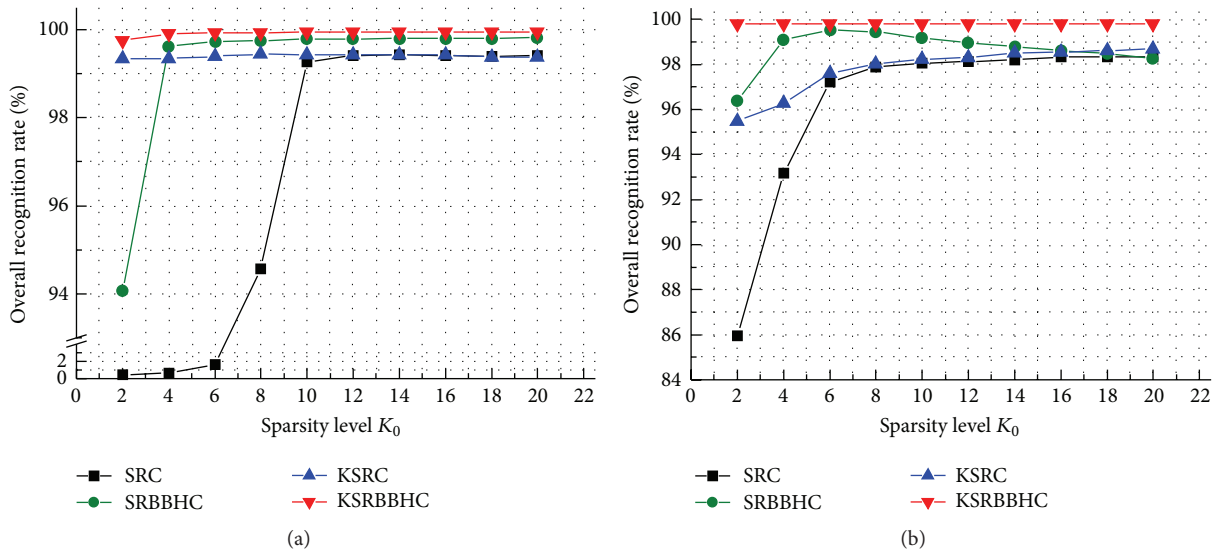


FIGURE 3: Classification accuracies of four sparse-based classifiers with different  $K_0$  for two datasets: (a) ROI-I; (b) ROI-II when  $\sigma$  is optimized.

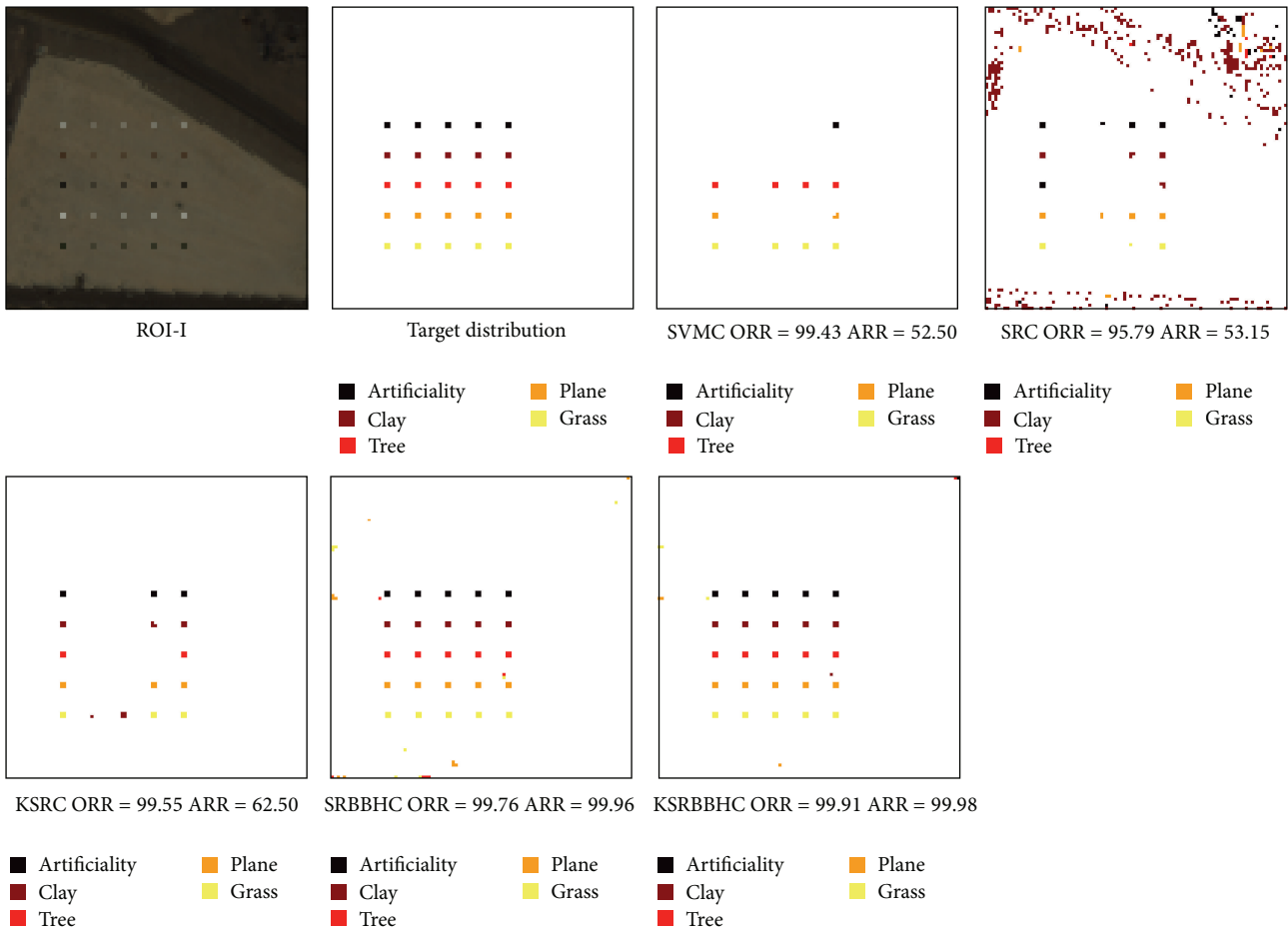


FIGURE 4: Classification maps of classifiers for ROI-I.



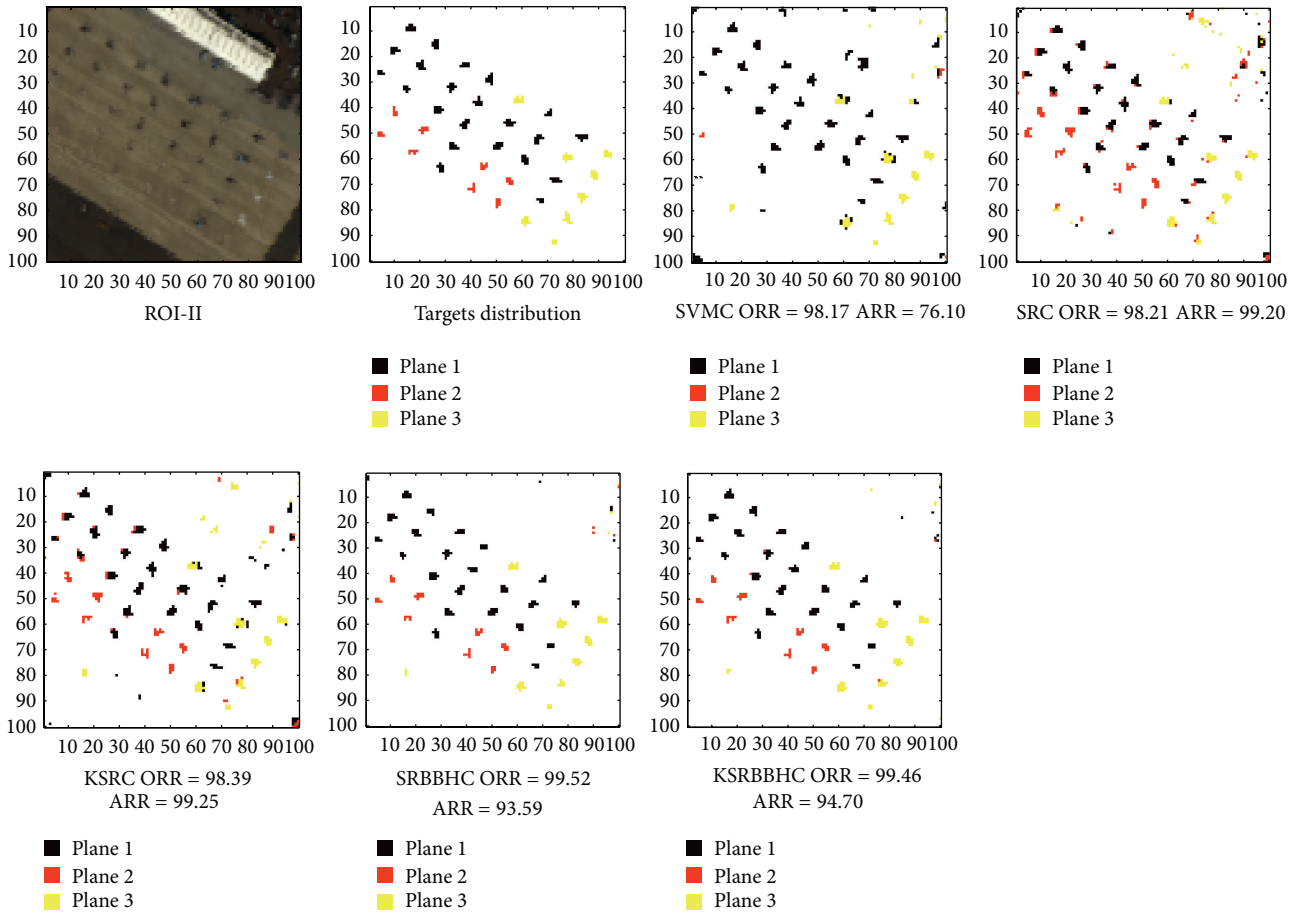


FIGURE 5: Classification maps of classifiers for ROI-II.

nonlinearly separable and complex data structure into consideration, the kernel version of SRBBHC (KSRBBHC) is then proposed to solve the complicated classification problem.

The proposed algorithms are tested on two HSI datasets, and the experimental results confirm the effectiveness of the proposed SRBBHC and KSRBBHC. From the experiment results, it can be concluded that (1) by collecting desired background training samples adaptively, the proposed local classifier solves the small target classification problem without requiring the number of classes present in image scene; (2) the SRBBH model enhances the discriminative power, which results in a better classification performance; (3) the kernelization further improves the classification performance to a certain extent.

All bands are involved in the proposed method, while the HSI may have more discriminative power at specific spectral channels. And the performance of SRBBHC and KSRBBHC is heavily influenced by background dictionary. Thereby, the performance can be further improved by some preprocess such as band selection and dictionary learning. The effect of spatial information [21] on the performance of SRBBHC and KSRBBHC will be investigated in our future work. In addition, we will also investigate the selection of appropriate validation threshold. By the way, backanalyzing

the distribution of vector  $\mathbf{R}$  of desired pixels and undesired pixels is a potential direction.

### Competing Interests

The authors declare that they have no competing interests.

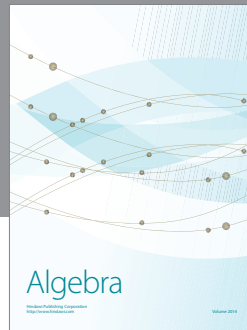
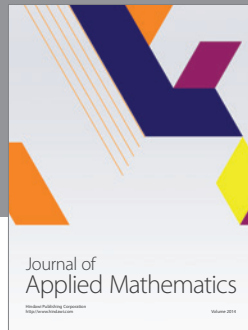
### Acknowledgments

The authors would like to thank the Wuhan University, Wuhan, China, for providing the San Diego airport data used in the experiments. This work is supported by the National Natural Science Foundation of China under Grant 61273275 and Aeronautical Science Foundation of China under Grant 20130196004.

### References

- [1] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [2] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the 5th Annual*

- ACM Workshop on Computational Learning Theory (COLT '92)*, pp. 144–152, Pittsburgh, Pa, USA, July 1992.
- [3] J. A. Gualtieri and R. F. Cromp, “Support vector machines for hyperspectral remote sensing classification,” in *27th AIPR Workshop: Advances in Computer-Assisted Recognition*, vol. 3584 of *Proceedings of SPIE*, pp. 221–232, Washington, DC, USA, October 1998.
- [4] G. Camps-Valls and L. Bruzzone, “Kernel-based methods for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 6, pp. 1351–1362, 2005.
- [5] L. Bruzzone, M. Chi, and M. Marconcini, “A novel transductive SVM for semisupervised classification of remote-sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 11, pp. 3363–3372, 2006.
- [6] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [7] Z. Yuan, H. Sun, K. Ji, Z. Li, and H. Zou, “Local sparsity divergence for hyperspectral anomaly detection,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1697–1701, 2014.
- [8] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Hyperspectral image classification via kernel sparse representation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 217–231, 2013.
- [9] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Sparse representation for target detection in hyperspectral imagery,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 629–640, 2011.
- [10] W. Liguó and Z. Chunhui, *Processing Techniques of Hyperspectral Imagery*, National Defense Industry Press, Beijing, China, 2013 (Chinese).
- [11] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Hyperspectral image classification using dictionary-based sparse representation,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, 2011.
- [12] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Simultaneous joint sparsity model for target detection in hyperspectral imagery,” *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 4, pp. 676–680, 2011.
- [13] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [14] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [15] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, New York, NY, USA, 2010.
- [16] H. Kwon and N. M. Nasrabadi, “Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4099–4109, 2010.
- [17] V. M. Patel and R. Chellappa, *Sparse Representations and Compressive Sensing for Imaging and Vision*, Springer Briefs in Electrical and Computer Engineering, Springer, New York, NY, USA, 2013.
- [18] Y. Zhang, L. Zhang, and S. Wang, “A Nonlinear sparse representation-based binary hypothesis model for hyperspectral target detection,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2513–2522, 2014.
- [19] T. Wang, B. Du, and L. Zhang, “A kernel-based target-constrained interference-minimized filter for hyperspectral sub-pixel target detection,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 626–637, 2013.
- [20] Y. Tang, S. Huang, Q. Ling, and Y. Zhong, “Adaptive kernel collaborative representation anomaly detection for hyperspectral imagery,” *High Power Laser and Particle Beams*, vol. 27, no. 9, Article ID 091008, 2015.
- [21] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Multiple spectral-spatial classification approach for hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4122–4132, 2010.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

