Chinese Journal of Electronics
Vol.23, No.1, Jan. 2014

A Novel Supervised Method for Hyperspectral Image Classification with Spectral-Spatial Constraints*

SUN Le, WU Zebin, LIU Jianjun and WEI Zhihui

(*Nanjing University of Science and Technology, Nanjing 210094, China*)

Abstract — In this paper, a new supervised classification method, combining spectral and spatial information, is proposed. The method is based on the two following facts. First, a hyperspectral pixel can be sparsely represented by a linear combination of the dictionary consists of a few labeled samples. If any unknown hyperspectral pixel lies in the subspace spanned by some labeled-class samples, it will be classified to this labeled-class. And this is to solve a fully constrained sparse unmixing problem with the l_2 regularization and the criterion of classification is relaxed to be determined by the largest value of sparse vector whose nonzero entries correspond to the weights of the labeled samples. Second, since the nearest neighbors probably belong to the same class, a spatial constraint is introduced. Alternating direction method of multipliers (ADMM) and the graph cut based method are then used to solve the spectral-spatial model. Finally, two real hyperspectral data sets are used to validate our proposed method. Experimental results show that the proposed method outperforms many of the state-of-the-art methods.

Key words — Hyperspectral classification(HC), Fully constrained sparse unmixing, Spatial constraint, Alternating direction method of multipliers (ADMM), Graph cut.

I. Introduction

Hyperspectral remote sensors capture digital images in hundreds of continuous narrow (about 2 to 10nm) spectral bands spanning the visible to infrared spectrum (400nm–2500nm)[1]. Pixels in Hyperspectral imaging (HSI) are represented by vectors whose entries corresponding to the spectral bands. Different materials usually reflect electromagnetic energy differently at specific wavelengths. And this allows the characterization, identification, and classification of the land-covers with improved accuracy and robustness. In recent years, many techniques have been developed for HSI classification. Among these methods, Support vector machines (SVMs) have been a powerful tool to solve supervised classification problems for high-dimensional data and have shown a good performance for hyperspectral classification[2−4], due to their ability to deal with large input spaces efficiently and produce sparse solutions. Another efficient methods are graph based methods[5,6], in which each sample spreads its labeled and unlabeled samples until a global stable state is achieved on the hole dataset. This kind of methods play a key role in hyperspectral classification. In addition, sparse multinomial logistic regression methods[7,8] based on Bayesian learning framwork also provide good performances and draw more attention in hyperspectral classification. This kind of methods learn the class distribution themselves and provide a sparse regressor. Recently, sparse representation has also been proposed for HSI classification[9,10]. It relies on the assumption that hyperspectral pixels in the same class lie in the same low-dimensional subspace. Thus, an unknown pixel can be sparsely represented by a set of training samples in a dictionary, and it doesn't need to learn the dictionary but use the training samples as the dictionary. In addition, a trend of hyperspectral classification is to include the spatial information[11,12], as well as to use the kernel method[13−15] to project the data into a nonlinear feature space, for improving the classification accuracy.

In this paper, a new supervised classification method with spectral-spatial constraints is proposed. we use the sparse unmixing to do the first step of classification and then impose the spatial-contextual information, which encourages the neighboring samples to belong to the same class, to improve the classification accuracy (see Fig.1). In the sparse unmixing procedure, the labeled samples are assumed to be spectrally pure and used as endmembers. And the corresponding fractional abundances (sparse vectors), imposing two constrains, sum-to-one (the so-called Abundance sum constraint–ASC) and non-negative (Abundance nonnegativity constraint–ANC), are estimated by the fully-constrained l_2-SUADMM (l_2 regularized Spectral unmixing by alternating direction method of multipliers[16]) method. The first step of classification is not determined by the minimum residual but relaxed to be determined by the maximum element of abundance, which is got under the two constraints. In order to improve the classification accuracy, the spatial-contextual information is proposed by forcing constraint on the classified neighbors instead of the sparse vectors. The main novelty of our proposed work is the integration of l_2 regularized sparse unmixing method with the spatial-contextual term forced on the classified neighbors.

II. HSI Classification Based on Sparse Representation

For the convenience of describing hyperspectral classification

*Manuscript Received Dec. 2012; Accepted Dec. 2012. This work is supported by Jiangsu Provincial Natural Science Foundation of China (No.BK2011701), the National Natural Science Foundation of China (No.61101194), Research Fund for the Doctoral Program of Higher Education of China (No.20113219120024), Project of China Geological Survey (No.1212011120227) and Jiangsu Innovation Projects (No.CXZZ12-0206).

problem, we define some notations. Let $\mathcal{S} \equiv \{1, \cdots, N\}$ be a set of integers indexing N pixels in a HSI, $\mathcal{L} \equiv \{1, \cdots, K\}$ be a set of K class labels, $\boldsymbol{X} \equiv (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N) \in \mathbb{R}^{L \times N}$ be a HSI with N pixels of L-dimension, where $\boldsymbol{x}_i$ is a L-dimension vector corresponding to pixel i, $\boldsymbol{y} \equiv (y_1, y_2, \cdots, y_N) \in \mathcal{L}^N$ be an image of class labels and $\boldsymbol{A}^k \equiv (\boldsymbol{a}_1^k, \boldsymbol{a}_2^k, \cdots, \boldsymbol{a}_{l^k}^k) \in \mathbb{R}^{L \times l^k}$ be a sub-dictionary containing l^k labeled samples in class k.

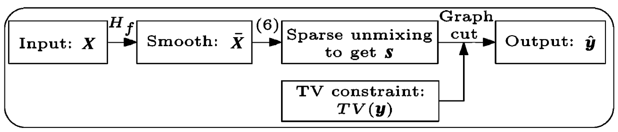


Fig. 1. Block diagram summarizing the most relevant steps of our proposed l_2-SUADMM-TV algorithm

1. Sparse unmixing model for HSI

The sparse model is based on the assumption that hyperspectral pixels in the same class approximately lie in the same low dimensional subspace. It is to say that, if an unlabeled hyperspectral pixel $\boldsymbol{x} \in \boldsymbol{X}$ belongs to the kth class, its spectrum turns to lie in a low-dimensional subspace spanned by the labeled hyperspectral samples in class k, and can be represented by a linear combination of the labeled samples, as[9]:

$$\begin{aligned} \boldsymbol{x} &\approx s_1^k \boldsymbol{a}_1^k + s_2^k \boldsymbol{a}_2^k + \cdots + s_{l^k}^k \boldsymbol{a}_{l^k}^k \\ &= \underbrace{[\boldsymbol{a}_1^k \boldsymbol{a}_2^k \cdots \boldsymbol{a}_{l^k}^k]}_{\boldsymbol{A}^k} \cdot \underbrace{[s_1^k s_2^k \cdots s_{l^k}^k]^T}_{\boldsymbol{s}^k} \\ &= \boldsymbol{A}^k \cdot \boldsymbol{s}^k \end{aligned} \tag{1}$$

where $\boldsymbol{s}^k$ is a l^k-dimensional vector whose entries are the abundances of the corresponding spectra in $\boldsymbol{A}^k$.

Under the above assumption, if we known the class number in the hyperspectral scene, any test samples would lie in one of the subspace spanned by one class of all K classes. Then, combining the dictionaries $\{\boldsymbol{A}^k\}_{k=1,2,\cdots,K}$, a unlabeled sample $\boldsymbol{x}$ can be written as a linear combination of all labeled samples as:

$$\begin{aligned} \boldsymbol{x} &\approx \boldsymbol{A}^1 \boldsymbol{s}^1 + \boldsymbol{A}^2 \boldsymbol{s}^2 + \cdots + \boldsymbol{A}^K \boldsymbol{s}^K \\ &= \underbrace{[\boldsymbol{A}^1 \cdots \boldsymbol{A}^K]}_{\boldsymbol{A}} \cdot \underbrace{\begin{bmatrix} \boldsymbol{s}^1 \\ \vdots \\ \boldsymbol{s}^K \end{bmatrix}}_{\boldsymbol{s}} \\ &= \boldsymbol{A} \cdot \boldsymbol{s} \end{aligned} \tag{2}$$

where $\boldsymbol{A}$ is a $L \times l$ matrix containing all labeled samples from K classes with $l = \sum_{k=1}^{K} l^k$, and the vector $\boldsymbol{s}$ is sparse ($\boldsymbol{s}$ contains a few non-zeros entries).

In contrast with the ideal case, there are always noise in the observed hyperspectral pixel $\boldsymbol{x}$. In order to solve the reconstruction problem imposing noise and find the sparse vector $\boldsymbol{s}$ for a unlabeled sample, it's necessary to impose a sparsity regularization for the problem, as the noise will be amplified by the eigenvalues in the Singular value decomposition (SVD) of the matrix $\boldsymbol{A}$ at its bad-condition (the condition number of $\boldsymbol{A}$ is big) case. One of the choice of the sparsity regularization is the l_0 norm of $\boldsymbol{s}$, which gives the number of non-zero elements of $\boldsymbol{s}$, and then the sparse vector $\boldsymbol{s}$ can be found by the following optimization problem:

$$\begin{aligned} &\boldsymbol{s} = \arg\min_{\boldsymbol{s}} ||\boldsymbol{s}||_0 \\ &\text{s.t} \quad ||\boldsymbol{A} \cdot \boldsymbol{s} - \boldsymbol{x}||_2 < \varepsilon \end{aligned} \tag{3}$$

But, this problem is NP-hard and can only be approximately solved by greedy pursuit algorithms[17−19]. In this paper, we turn to solve a easy and convex problem using the l_1 regularization of the weighted vector $\boldsymbol{s}$, the problem becomes:

$$\begin{aligned} &\boldsymbol{s} = \arg\min_{\boldsymbol{s}} ||\boldsymbol{s}||_1 \\ &\text{s.t} \quad ||\boldsymbol{A} \cdot \boldsymbol{s} - \boldsymbol{x}||_2 < \varepsilon \end{aligned} \tag{4}$$

As we assume that the labeled samples can be seen as the endmembers in the hyperspectral scene, the problem is a sparse unmixing problem[20−23]. Imposing the two constraints ANC and ASC and using the Lagrange multiplier method representation, the problem is formulated as:

$$\begin{aligned} &\boldsymbol{s} = \arg\min_{\boldsymbol{s}} \left\{ \frac{1}{2} ||\boldsymbol{A} \cdot \boldsymbol{s} - \boldsymbol{x}||_2^2 + \lambda \cdot ||\boldsymbol{s}||_1 \right\} \\ &\text{s.t} \quad \boldsymbol{s} \geq 0, \quad \mathbf{1}^T \cdot \boldsymbol{s} = 1 \end{aligned} \tag{5}$$

where $\mathbf{1}^T \cdot \boldsymbol{s} = 1$ denotes the sum of elements of $\boldsymbol{s}$ is 1. Eq.(5) is the traditional $l_2 - l_1$ problem, and can be solved by existing methods[24,25]. As $||\boldsymbol{s}||_1 \equiv 1$ under the ASC constraint, we will use the $||\boldsymbol{s}||_2^2$ to replace it in the following section.

2. Proposed sparse unmixing method by minimizing l_2 norm of abundance vector

In this section, we use the l_2 norm instead of the l_1 norm to obtain the sparse solution under the ANC and ASC constraints. The advantages of using l_2 norm is that (1) under the ASC constraint, l_1 constraint on $\boldsymbol{s}$ is always equal to 1, (2) l_2 norm is easier than l_1 as well as that, under the ASC constraint, maximizing $||\boldsymbol{s}||_2$ (defined as $||\boldsymbol{s}||_2 = \sqrt{\sum_i |s_i|^2}$, where s_i is the ith element of $\boldsymbol{s}$) can also get a sparse solution, sometimes, sparser than that got by minimizing $||\boldsymbol{s}||_1$. For example, given an unlabeled sample $\boldsymbol{x}1$ and the labeled samples matrix $\boldsymbol{A}$, if $\boldsymbol{s}1 = (0.2, 0, 0.3, 0, 0.5)^T$ and $\boldsymbol{s}2 = (0.1, 0, 0, 0.9, 0)^T$ are two possible solution of $\{\arg\min_{\boldsymbol{s}} ||\boldsymbol{x}1 - \boldsymbol{A} \cdot \boldsymbol{s}||_2^2\}$ under the constraints (ANC and ASC). In this situation, we will get the sparser solution $\boldsymbol{s}2$ by maximizing $||\boldsymbol{s}||_2$ regularization. And the sparsity of maximizing $\boldsymbol{s}2$ and minimizing $\boldsymbol{s}1$ are shown in Fig.2. So, we suggest using the l_2 norm instead of the l_1 norm.

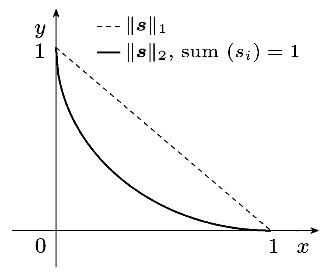


Fig. 2. The sparsity that may be reached by maximizing $||\boldsymbol{s}||_2$ regularization under ASC and minimizing $||\boldsymbol{s}||_1$ regularization separately

The new model is formulated as:

$$\begin{aligned} &\boldsymbol{s} = \arg\min_{\boldsymbol{s}} \left\{ \frac{1}{2} ||\boldsymbol{A} \cdot \boldsymbol{s} - \boldsymbol{x}||_2^2 - \frac{\lambda}{2} \cdot ||\boldsymbol{s}||_2^2 \right\} \\ &\text{s.t} \quad \boldsymbol{s} \geq 0, \quad \mathbf{1}^T \cdot \boldsymbol{s} = 1 \end{aligned} \tag{6}$$

this is a constrained quadratic problem and easy to solve. Here, we use the ADMM to solve it, the reader can refer to Ref.[16] for the details about ADMM.

Now, we transfer the constrained model into an unconstrained model using Lagrange multipliers:

$$\boldsymbol{s} = \arg\min_{\boldsymbol{s}} \left\{ \frac{1}{2} ||\boldsymbol{A} \cdot \boldsymbol{s} - \boldsymbol{x}||_2^2 - \frac{\lambda}{2} \cdot ||\boldsymbol{s}||_2^2 + l_{\{1\}}(\mathbf{1}^T \cdot \boldsymbol{s}) + l_{R_+^n}(\boldsymbol{s}) \right\} \tag{7}$$

where R_+^n means the positive vector space and $l_x(\cdot)$ is the indicator function, which means that if $(\cdot)$ belongs to the set x, the function equals to 1, else $+\infty$. According to the ADMM method, we get the following iterative formulation:

$$\begin{cases} \boldsymbol{s}_{k+1} \in \arg\min_{\boldsymbol{s}} \left\{ \frac{1}{2} ||\boldsymbol{x} - \boldsymbol{A} \cdot \boldsymbol{s}||_2^2 - \frac{\lambda}{2} \cdot ||\boldsymbol{s}||_2^2 + \frac{\mu}{2} ||\boldsymbol{s} - \boldsymbol{u}_k - \boldsymbol{d}_k||_2^2 + l_{\{1\}}(\mathbf{1}^T \cdot \boldsymbol{s}) \right\} & (8) \\ \boldsymbol{u}_{k+1} \in \arg\min_{\boldsymbol{u}} \left\{ l_{R_+^n}(\boldsymbol{u}) + \frac{\mu}{2} ||\boldsymbol{s}_{k+1} - \boldsymbol{u} - \boldsymbol{d}_k||_2^2 \right\} & (9) \\ \boldsymbol{d}_{k+1} \leftarrow \boldsymbol{d}_k - (\boldsymbol{s}_{k+1} - \boldsymbol{u}_{k+1}) & (10) \end{cases}$$

Eq.(8) is the standard equality constrained quadratic problem and can be formulated as follows.

$$\boldsymbol{s}_{k+1} \in \arg\min_{\boldsymbol{s}} \left\{ \frac{1}{2}\boldsymbol{s}^T(\boldsymbol{A}^T\boldsymbol{A} + (\mu-\lambda)\boldsymbol{I})\boldsymbol{s} - (\boldsymbol{A}^T\boldsymbol{x} + \mu(\boldsymbol{u}_k + \boldsymbol{d}_k))^T\boldsymbol{s} \right\} \quad \text{s.t} \quad \mathbf{1}^T \cdot \boldsymbol{s} = 1 \tag{11}$$

the solution of Eq.(11) is:

$$\boldsymbol{s}_{k+1} \leftarrow \boldsymbol{H}^{-1}\boldsymbol{w} - \boldsymbol{c}\mathbf{1}^T\boldsymbol{H}^{-1}\boldsymbol{w} + \boldsymbol{c}$$

where

$$\begin{aligned} \boldsymbol{H} &= \boldsymbol{A}^T\boldsymbol{A} + (\mu-\lambda)\boldsymbol{I} \\ \boldsymbol{w} &= \boldsymbol{A}^T\boldsymbol{x} + \mu(\boldsymbol{u}_k + \boldsymbol{d}_k) \\ \boldsymbol{c} &= \boldsymbol{H}^{-1}\mathbf{1}(\mathbf{1}^T\boldsymbol{H}^{-1}\mathbf{1})^{-1} \end{aligned}$$

and $\mathbf{1}$ is the column vector with elements of ones. Eq.(9) is an inequality constrained quadratic problem and can also be easily solved, the solution is formulated as:

$$\boldsymbol{u}_{k+1} = P_+(\boldsymbol{s}_{k+1} - \boldsymbol{d}_k) \tag{12}$$

where P_+ is a projection operator for the first orthant. $\boldsymbol{d}_{k+1}$ in Eq.(10) is the Lagrange multipliers.

In order to reduce the impact of the spectral noise and enhance the spatial piecewise smoothness in some sense, we use a filter function to convolute the pixel $\boldsymbol{x}$ and rewrite Eq.(6) as:

$$\boldsymbol{s} = \arg\min_{\boldsymbol{s}} \left\{ \frac{1}{2}||\boldsymbol{A}\cdot\boldsymbol{s} - H_f * \boldsymbol{x}||_2^2 - \frac{\lambda}{2}\cdot||\boldsymbol{s}||_2^2 \right\} \quad \text{s.t} \quad \boldsymbol{s} \geq 0, \quad \mathbf{1}^T\cdot\boldsymbol{s} = 1 \tag{13}$$

where H_f is the filter (average filter or gaussian filter) function, it can be also seen as a kind of preprocessing. But, in the course of preprocessing, all labeled samples are not changed.

Algorithm 1 l_2-SUADMM sparse unmixing algorithm

- **Input** iteration number $k = 0$, maximum iteration number K, $\mu > 0$, $\boldsymbol{u}_0 = (0, 0, \cdots, 0)$ and $\boldsymbol{d}_0 = (0, 0, \cdots, 0)$
- **while** $k < K$ or other criterion is not satisfied **do**
- $\boldsymbol{w} = \boldsymbol{A}^T(H_f * \boldsymbol{x}) + \mu(\boldsymbol{u}_k + \boldsymbol{d}_k)$
- $\boldsymbol{s}_{k+1} \leftarrow \boldsymbol{H}^{-1}\boldsymbol{w} - \boldsymbol{c}\mathbf{1}^T\boldsymbol{H}^{-1}\boldsymbol{w} + \boldsymbol{c}$
- $\boldsymbol{u}_{k+1} \leftarrow \max(\boldsymbol{s}_{k+1} + \boldsymbol{d}_k, 0)$
- $\boldsymbol{d}_{k+1} \leftarrow \boldsymbol{d}_k - (\boldsymbol{s}_{k+1} - \boldsymbol{u}_{k+1})$
- $k \leftarrow k + 1$
- **end while**
- **Output** $\boldsymbol{s}$

Algorithm 1 provides a pseudocode for our proposed sparse unmixing method by maximizing the l_2 regularization under ANC and ASC. During the iterative procedure, the most computational complexity is calculating $\boldsymbol{H}^{-1}$. However, in hyperspectral classification, the labeled samples matrix $\boldsymbol{A}$ is fixed, we can precompute the SVD of $\boldsymbol{A}^T\boldsymbol{A}$, that is, if $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{S}\cdot\boldsymbol{\Sigma}\cdot\boldsymbol{V}$, then $\boldsymbol{H}^{-1} = \boldsymbol{S}\cdot\text{diag}(1./(\text{diag}(\boldsymbol{\Sigma}) + (\mu-\lambda)))\cdot\boldsymbol{V}$, where diag($\boldsymbol{x}$) denotes getting the main diagonal of $\boldsymbol{x}$, if $\boldsymbol{x}$ is a matrix or returning a matrix with $\boldsymbol{x}$ as its main diagonal, if $\boldsymbol{x}$ is a vector. Up to a constant, this inverse takes the same amount of time as multiplication. Furthermore, the parameter λ is suggested to be chosen as a small value to ensure the stability of the result.

3. Relaxed criterion for HSI classification

In this section, we relax the classification criterion from minimum residual to the maximum entries of the sparse vector $\boldsymbol{s}$ under the two constraints (ANC and ASC). First, we use the variables defined in Ref.[9], the kth residual (*i.e.*, error between the test sample $\boldsymbol{x}_i$ and the one reconstructed from labeled samples in the kth class) is defined as:

$$r^k(\boldsymbol{x}_i) = ||\boldsymbol{x}_i - \boldsymbol{A}^k\cdot\hat{\boldsymbol{s}}^k||_2, \quad i\in\mathcal{S}, \quad k\in\mathcal{L} \tag{14}$$

where $\hat{\boldsymbol{s}}^k$ is the estimation of the sparse vector $\boldsymbol{s}^k$, then, the class image can be obtained from the following formulation:

$$y_i \equiv class(\boldsymbol{x}_i) = \arg\min_{k\in\mathcal{L}}\{r^k(\boldsymbol{x}_i)\}, \qquad i\in\mathcal{S} \tag{15}$$

In order to simplify the classification criterion only using the sparse vector $\boldsymbol{s}$ instead of calculating the kth residual for $k = 1, 2, \cdots, K$, we note $\tilde{s}^k = \sum_{i=1}^{l^k}\hat{s}_i^k$ and $\varphi(y_i|\boldsymbol{A}^k) = \tilde{s}^k$, then the classification criteria can be simply written as:

$$y_i \equiv class(\boldsymbol{x}_i) = \arg\max_{k\in\mathcal{L}}\{\varphi(y_i|\boldsymbol{A}^k)\}, \qquad i\in\mathcal{S} \tag{16}$$

where $\tilde{s}^k$ is a value not vector, and it is the sum of the entries of vector $\hat{\boldsymbol{s}}^k$.

As columns of the labeled samples matrix $\boldsymbol{A}$ are not normalized, it is necessary to impose the ASC constraint to relax the criterion to depend on $\boldsymbol{s}$ only. This is the reason why we have to force the ASC constraint during the sparse unmixing procedure. Under the ASC and ANC constraints, we have $\sum_{k=1}^{K}\varphi(y_i|\boldsymbol{A}^k) = 1$, so $\varphi(y_i|\boldsymbol{A}^k)$ can be seen as a probability of y_i belonging to the kth class in some sense. Hence, the relaxation of classification criterion from Eq.(15) to Eq.(16) is reasonable.

III. Proposed Model Imposing Spatial Constraint

The classification criterion Eq.(16) is only focus on the spectral sparse representation aspect without considering the spatial information, which consequently leads to the low accuracy of classification. In order to encourage the spatial information, we add a regularization term to the optimal problem as follows:

$$\text{TV}(\boldsymbol{y}) = \sum_{|i-j|<\delta}|y_i - y_j| \tag{17}$$

where $y_i \in \mathcal{L}$ for each $i \in \mathcal{S}$, and $|\cdot|$ denotes the absolute value, $|i-j| < \delta$ denotes the ith pixel and the jth pixel are neighbors and δ controls the size of neighborhood. This TV regularizer encourages the pixels in the same neighborhood to belong to the same class and it gives no preference to any direction. So this regularization term would promote piecewise smooth classifications.

In summary, we have used the linear sparse regression to do the classification from the spectral viewpoint and then added a TV-induced regularization to imposing the spatial information according to the prior that hyperspectral image turn to be piecewise in spatial domain. Combining the two aspects, our spatial-spectral classification model can be written as:

$$\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}\in\mathcal{L}}\left\{-\sum_{i\in\mathcal{S}}\varphi(y_i|\boldsymbol{A}^k) + \mu_s\cdot\sum_{|i-j|<\delta}|y_i - y_j|\right\} \tag{18}$$

the first term is spectral term imposing the spectral information, which is solved by the fully constrained l_2 sparse representation method. The second term is the TV term which encourages the pixels in neighborhood to belong to the same class in spatial domain. μ_s is the smoothness parameter balancing the spectral term and spatial term. It is difficult to solve this TV regularized classification model for the discrete status of y_i. In order to solve this optimization problem of Eq.(18), we turn to graph cuts[26−29] recently developed energy minimization algorithms, which are efficient tools to tackle this kind of optimization problems. The relationship associating our model with graphical model is presented in Fig.3. We use the graph cut algorithm proposed in Ref.[29] to solve optimization problem Eq.(18) for its polynomial computation.

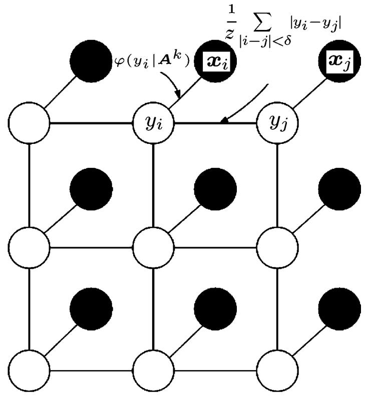


Fig. 3. Graphical example on hyperspectral classification

Algorithm 2 provides a pseudocode for our proposed supervised classification algorithm combining sparse unmixing and TV-induced spatial constraint. This algorithm, termed as l_2-SUADMM-TV hereinafter, integrates all modules described in above two sections. Line 3 in Algorithm 2 learns the sparse vector by solving the fully constrained unmixing problem, and the filter function H_f could be set as a gaussian filter with deviation of 1 and the window size is 3×3.

Algorithm 2 l_2-SUADMM-TV classification

1. **Input**: class number K, labeled samples set $\boldsymbol{A} = [\boldsymbol{A}^1, \boldsymbol{A}^2, \cdots, \boldsymbol{A}^K]$, test sample $\boldsymbol{x}_i$, number of image pixels N, filter function H, smoothness parameter μ.
2. **while** $i < N$ **do**
3. $\hat{\boldsymbol{s}} = l_2\text{-SUADMM}(\boldsymbol{A}, H_f * \boldsymbol{x}_i)$
4. $\hat{\varphi}(y_i|\boldsymbol{A}^k) = \tilde{s}^k$
5. $\hat{y}_i = Graphcut(\hat{\varphi}(y_i|\boldsymbol{A}^k), \mu_s)$
6. $i = i + 1$
7. **end while**
8. **Output**: $\hat{\boldsymbol{y}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_N)$

IV. Experiments on Real Hyperspectral Data

In this section, we show the effectiveness of the proposed method using two real hyperspectral data sets. For comparison, we adopt several state-of-the-art supervised classifiers such as Subspace pursuit with smoothness(SP-S)[10], Simultaneous orthogonal matching pursuit (SOMP)[10], Logistic regression via variable splitting and augmented lagrangian (LORSAL-MLL)[11] and SVMs[4,30], which are well-established techniques in the machine learning community. For SVM, we use a composite kernel (denoted by SVM-CK) that combines the spectral and spatial information via a weighted kernel summation, which has been shown to outperform the spectral-only SVM in HSI classification. All parameters of these methods were set according to the reference papers. And the parameters in our proposed method, if no other specifying, are set as follows: $\lambda = 0.00001$, $\mu = 0.01$, $\mu_s = 2$ and H_f is chosen as Gaussian filter with deviation of 1 and window size of 3×3.

The classification results are measured by the Overall accuracy (OA), Average accuracy (AA) and the k statistic. The OA is computed by the ratio between correctly classified test samples and the total number of test samples, and the AA is the mean of all class accuracies. The k statistic is computed by weighting the measured accuracies. It incorporates both of the diagonal and off-diagonal entries of the confusion matrix and is a robust measure of the degree of agreement. And all these measurements in the following experiments are achieved by the adopted classifiers after ten Monte Carlo runs.

1. AVIRIS Indian Pines data set

The first data set was collected by the Airborne visible/infrared imaging spectrometer (AVIRIS) sensor over the Indian Pines region in June 1992. The AVIRIS sensor generates 220 bands across the spectral range from 0.2μm to 2.4μm. In the experiments, the number of bands is reduced to 200 by removing 20 water absorption bands[30]. This image has spectral resolution of 10nm and spatial resolution of 20m by pixel, and the spatial dimension is 145×145. The ground truth contains 16 land cover classes and a total of 10366 labeled pixels. The number of pixels in the smallest class is 20, while the number of pixels in the largest class is 2468, as seen in Table 1.

Fig.4 and Fig.5 show the performance of our method at the different values of smoothness parameter μ_s and the training samples on Indian Pines data set, separately. From Fig.4, we conclude that the classification performance indeed depends on the setting of μ_s. However, the proposed method is not sensitive to the setting of μ_s, even with the poorest $\mu_s = 20$, it leads to a good classification result. From Fig.5, we conclude that the classification results produced by our method show high accuracy even with very limited training samples. As the number of labeled samples increases, the

Table 1. Classification accuracy (%) for the Indian Pines image on the test set

Class	Name	Train	Test	SVM	SVM-CK	SP-S	SOMP	l_1	LORSAL-MLL	l_2-SUADMM	l_2-SUADMM-TV
1	Alfalfa	6	48	81.25	95.83	87.50	85.42	39.58	72.08	92.71	**96.88**
2	Corn-notill	144	1290	86.28	96.67	91.94	94.88	78.53	93.53	96.01	**98.72**
3	Corn-min	84	750	72.80	90.93	82.53	94.93	51.87	90.05	91.07	**97.33**
4	Corn	24	210	58.10	85.71	70.95	91.43	28.57	**97.38**	89.52	93.81
5	Grass/Pasture	50	447	92.39	93.74	94.41	89.49	80.76	94.90	97.99	**98.43**
6	Grass/Trees	75	672	96.88	97.32	99.26	98.51	99.40	98.42	99.93	**100**
7	Grass/Pasmowed	3	23	43.48	69.57	47.83	**91.30**	17.39	55.22	89.13	89.13
8	Hay-windtrowed	49	440	98.86	98.41	99.77	99.55	99.32	99.55	99.77	**100**
9	oats	2	18	50.00	55.56	**94.44**	0	16.67	42.48	55.56	55.56
10	Soybeans-notill	97	871	71.53	93.80	86.80	89.44	63.95	90.01	86.11	**95.41**
11	Soybeans-min	247	2221	84.38	94.37	93.38	97.34	86.04	96.94	97.52	**99.82**
12	Soybeans-clean	62	552	85.51	93.66	84.24	88.22	57.79	**97.79**	93.66	97.64
13	Wheat	22	190	**100**	99.47	**100**	**100**	**100**	99.68	**100**	**100**
14	Woods	130	1164	93.30	99.14	98.28	99.14	97.94	97.88	99.61	**99.91**
15	Bui-GrassTrees	38	342	64.91	87.43	69.30	**99.12**	35.96	87.05	84.94	93.42
16	Stone-steelTowers	10	85	88.24	**100**	95.29	96.47	90.59	84.24	94.12	**100**
Overall accuracy (%)				84.52	94.86	91.16	95.28	77.99	94.82	95.31	**98.40**
Average accuracy (%)				79.24	90.73	87.25	88.45	65.27	87.33	91.73	**94.75**
k statistic				0.823	0.941	0.899	0.946	0.746	0.941	0.9464	**0.9818**

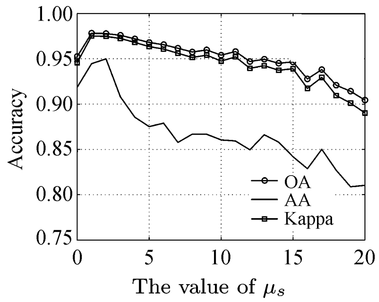


Fig. 4. The OA, AA and kappa statistic as a function of the smoothness parameter μ_s with training samples $l = 1043$

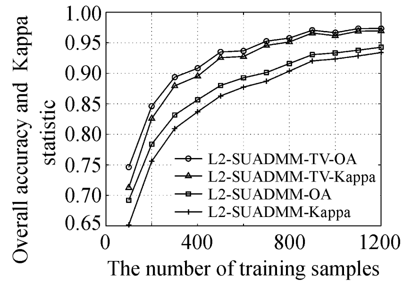


Fig. 5. The OA and kappa statistic (k) as a function of the number of training samples l

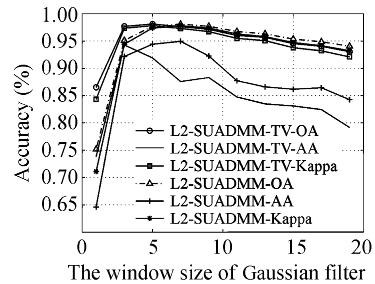


Fig. 6. The OA, AA and kappa statistic (k) as a function of the window size of filter H_f

OA and k statistic increase. Fig.6 shows the OA, AA and kappa statistic (k) as a function of the window size of Gaussian filter H_f. It is easy to see that our method is robust on the change of window sizes. Although combining the preprocessing (Gaussian filter with deviation of 1) and sparse unmixing still leads to a good result, our final method is still a little more accurate and robuster than it.

In order to test the performance of the proposed method with limited training sets, a total size of $l = 1043$ (which represents about 10% of the available labeled samples among classes) was used for training purpose and the training samples of each class was chosen randomly according to Table 1, where the remaining about 90% of the samples were used for validation. Table 1 illustrates the OA, AA, k statistic and the accuracy of each class for all tested methods. It is easy to see that the classifiers imposing spatial information produce better results, all more than 90% accuracy. And among these, our proposed method l_2-SUADMM-TV produces the best results in all OA, AA and k statistic. For illustrative purposes, Fig.7 shows the ground truth and some of the classification results obtained by the different tested classifiers for the Indian Pines Scene. For each classifier, we randomly selected one of the maps obtained after conducting ten Monte Carlo runs.

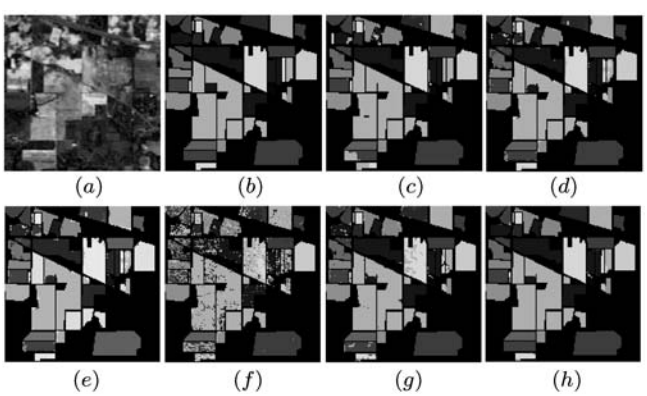


Fig. 7. AVIRIS image. (a) The gray map of Indian Pines; (b) Ground truth and the overall accuracy of the methods of (c) LORSAL-MLL (OA = 94.82%); (d) SVM-CK (OA =94.86%); (e) SOMP (OA = 95.28%); (f) l_1 (OA =76.59%); (g) l_2-SUADMM (OA = 94.98%) and (h) l_2-SUADMM-TV (OA = 98.75%) with about 10% training samples

2. ROSIS university of Pavia data set

The second real hyperspectral data set that used in our experiment was acquired in 2001 by the Reflective optics system imaging spectrometer (ROSIS), flown over the city of Pavia, Italy. The sensor generates 115 spectral bands ranging from 0.43μm to 0.86μm and has a spatial resolution of 1.3m per pixel. The image scene, with size of 610×340 pixels, is centered at the University of Pavia and has 103 bands after removing 12 noisiest bands. There are nine ground truth classes and 3921 (about 9%) of all labeled data are used as training and the rest are used for testing. We also adopted the above classifiers for comparison, the results are shown in Table 2.

From it, we can conclude the same results as that in Indian Pines experiment. Our method outperforms all of the other classifiers. Fig.8 illustrates the classification maps achieved by some of the considered methods.

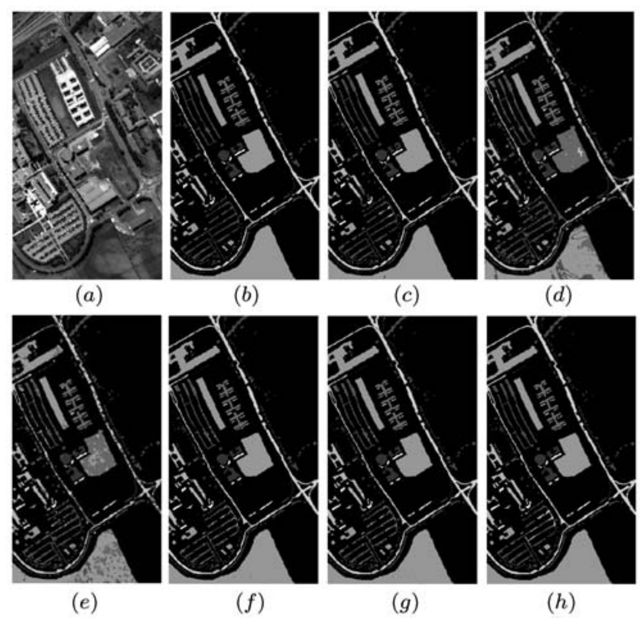


Fig. 8. ROSIS image. (a) The gray map of ROSIS image; (b) Ground truth and the overall accuracy of the methods of (c) LORSAL-MLL (OA = 97.08%); (d) SVM-CK (OA =87.18%); (e) SOMP (OA = 79%); (f) l_1 (OA =76.59%); (g) l_2-SUADMM (OA = 96.76%) and (h) l_2-SUADMM-TV (OA = 98.04%) with about 9% training samples

V. Conclusions and Future Work

In this paper, we have proposed a novel supervised classification method combining sparse unmixing and spatial information. In the proposed method, an HSI pixel is assumed to be sparsely represented by a few atoms consist of the training samples. In order to recover the sparse vector of a test spectral sample, we assume the training samples to be purely spectra (endmembers) and solve the

Table 2. Classification accuracy (%) for the university of Pavia image on the test set

Class	Name	Train	Test	SVM	SVM-CK	SP-S	SOMP	l_1	LORSAL-MLL	l_2-SUADMM	l_2-SUADMM-TV
1	Asphalt	548	6304	84.30	79.85	83.79	59.33	80.65	96.31	95.94	**98.03**
2	Meadows	540	18146	67.01	84.86	72.35	78.15	64.74	97.42	99.39	**99.92**
3	Gravel	392	1815	68.43	81.87	71.85	83.53	73.22	89.14	97.66	**99.94**
4	Trees	524	2912	97.80	96.36	98.94	96.91	98.35	98.35	**99.61**	97.56
5	Metal sheets	265	1113	99.37	99.37	**100.0**	99.46	99.91	99.85	99.81	99.81
6	Bare soil	532	4572	92.45	93.55	92.63	77.41	92.54	99.60	97.04	**100**
7	Bitumen	375	981	89.91	90.21	91.44	98.57	86.95	98.10	99.58	**99.69**
8	Bricks	514	3364	92.42	92.81	**95.57**	89.09	81.54	94.55	84.25	88.79
9	Shadows	231	795	97.23	95.35	98.24	91.95	98.99	**100**	**100**	99.86
Overall accuracy (%)				79.15	87.18	82.09	79.00	76.87	97.08	97.31	**98.56**
Average accuracy (%)				87.66	90.47	89.42	86.04	86.32	97.04	97.03	**98.18**
k statistic				0.37	0.833	0.772	0.728	0.709	0.960	0.9631	**0.980**

fully constrained unmixing problem by maximizing the l_2 norm of the sparse vector using the ADMM method, and give the reasons of using the l_2 regularization to replace the l_1 regularization as well as show the sparsity of l_2 regularization under ASC constraint. So as to ease of computation and the context consistency, we relax the classification criteria to only depend on the recovered sparse abundance vector. To improve the classification performance, we proposed a TV-induced spatial constraint to encourage the pixels in neighborhood to belong to the same class. At last, the spectra-spatial model is solved by graph cut based algorithm in polynomial computation. Experimental results on the real hyperspectral image show that the proposed algorithm yields high classification accuracy and outperforms most of the tested methods. In the future we would consider a more standardized spectral unmixing on (1) estimating the endmembers in the training samples and constructing the fractional abundances of each endmember of the scene, or (2) learning a dictionary (or designing a composite kernel containing spatial and spectral information) using the training samples in their feature space and using the learned dictionary for classification.

References

[1] J.B. Adams, M.O. Smith, A.R. Gillespie, "Imaging spectroscopy: Interpretation based on spectral mixture analysis", *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, Vol.7, pp.145–166, 1993.

[2] F. Melgani, 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.

[3] M. Fauvel, J.A. Benediktsson, J. Chanussot, J.R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.46, No.11, pp.3804–3814, 2008.

[4] A. Plaza, J.A. Benediktsson, J.W. Boardman, J. Brazile, L. Bruzzone and others, "Recent advances in techniques for hyperspectral image processing", *Remote Sensing of Environment*, Vol.113, pp.S110–S122, 2009.

[5] A. Blum, S. Chawla, "Learning from labeled and unlabeled data using graph mincuts", *Proceedings of the Eighteenth International Conference on Machine Learning*, San Francisco, CA, pp.19–26, 2001.

[6] G. Camps-Valls, T. Bandos Marsheva, D. Zhou, "Semi-supervised graph-based hyperspectral image classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.45, No.10, pp.3044–3054, 2007.

[7] J. Li, J.M. Bioucas-Dias, A. Plaza, "Hyperspectral image classification based on a fast bregman sparse multinomial logistic regression algorithm", *6th EARSeL SIG IS Workshop*, Tel-Aviv, Israel, pp.1–6, 2009.

[8] B. Krishnapuram, L. Carin, M.A.T. Figueiredo, A.J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.6, pp.957–968, 2005.

[9] Y. Chen, N.M. Nasrabadi, T.D. Tran, "Classification for hyperspectral imagery based on sparse representation", *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (*WHISPERS*), Reykjavik, Iceland, pp.1–4, 2010.

[10] Y. Chen, N.M. Nasrabadi, 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.

[11] J. Li, J.M. Bioucas-Dias, A. Plaza, "Exploiting spatial information in semi-supervised hyperspectral image segmentation", *IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (*WHISPERS*),Reykjavik, Iceland, pp.1–4, 2010.

[12] J. Li, J.M. Bioucas-Dias, A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.49, No.10, pp.3947–3960, 2011.

[13] G. Camps-Valls, L. Bruzzone, "Kernel-based methods for hyperspectral image classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.43, No.6, pp.1351–1362, 2005.

[14] Y. Chen, N.M. Nasrabadi, T.D. Tran, "Hyperspectral image classification via kernel sparse representation", *IEEE 18th International Conference on Image Processing* (*ICIP*), Brussels, Belguim, pp.1233–1236, 2011.

[15] C. Zhao, Y. Wang, F. Mei, "Kernel ICA feature extraction for anomaly detection in hypersepctral imagery", *Chinese Journal of Electronics*, Vol.21, No.2, pp.265–269, 2012.

[16] E. Esser, "Applications of Lagrangian-based alternating direction methods and connections to split Bregman", *CAM Report*, Vol.9, pp.1–31, 2009.

[17] S. Mallat, Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries", *IEEE Transactions on Signal Processing*, Vol.41, No.12, pp.3397–3415, 1993.

[18] J. Tropp and A. Gilbert, "Signal Recovery from Random Measurements via Orthogonal Matching Pursuit", *IEEE Transactions on Information Theory*, Vol.53, No.12, pp.4655–4666, 2007.

[19] 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.

[20] M.D. Iordache, A. Plaza, J. Bioucas-Dias, "On the Use of Spectral Libraries to perform Sparse Unmixing of Hyperspectral Data", *2nd IEEE GRSSWorkshop on Hyperspectral Image and Signal Processing WHISPERS*, Vol.42, No.5, pp.577–685, 2010.

[21] J.M. Bioucas-Dias, M.A.T. Figueiredo, "Alternating direction algorithm for constrained sparse regression: Application to hyperspectral unmixing", *IEEE Transactions on Image processing*, Vol.10, No.1, pp.266–277, 2010.

[22] M.D. Iordache, J. Bioucas-Dias, A. Plaza, "Umixing sparse hyperspectral mixtures", *International Journal of Computer Vision*, Vol.22, No.1, pp.61–79, 2009.

[23] M.D. Iordache, A. Plaza, "Recent developements in sparse hypespectral unmixing", *IEEE Geoscience and Remote Sensing Symposium* (*IGARSS*), Hawaii, USA, pp.1281–1284, 2010.

[24] J.M. Bioucas-Dias, M.A.T. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration", *IEEE Transactions on Image Processing*, Vol.16, No.12, pp.2992–3004, 2007.

[25] A. Szlam, Z. Guo, S. Osher, "A split bregman method for non-negative sparsity penalized least squares with applications to hyperspectral demixing", *IEEE 17th International Conference on Image Processing* (*ICIP*), HongKong, China, pp.1917–1920, 2010.

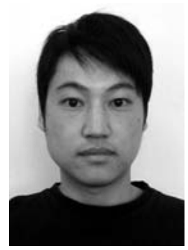
[26] C. Li, C.Y. Kao, J.C. Gore, Z. Ding, "Implicit active contours driven by local binary fitting energy", *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Minneapolis, Minnesota, USA, pp.1–7, 2007.

[27] Y. Boykov, V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.26, No.9, pp.1124–1137, 2004.

[28] Y. Boykov, O. Veksler, R. Zabih, "Fast approximate energy minimization via graph cuts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.23, No.11, pp.1222–1239, 2001.

[29] M. Schmidt, K. Alahari, "Generalized fast approximate energy minimization via graph cuts: Alpha-expansion beta-shrink moves", Available at: *http://hal.archives-ouvertes.fr/inria-00617524/*, 2011.

[30] J.A. Gualtieri, R.F. Cromp, "Support vector machines for hyperspectral remote sensing classification", *Proceedings of SPIE*, Vol.3584, pp.221–232, 1999.



SUN Le was born in Jiangsu Province in 1987. He received the B.S. degree from the School of Science, Nanjing University of Science and Technology, in 2009. He is now a Ph.D. student in the School of Computer Science and Engineering, Nanjing University of Science and Technology. His research interest is hyperspectral image processing. (Email: sunlecncom@163.com)

WU Zebin (corresponding author) was born in Zhejiang Province in 1981. He received the B.S. and Ph.D. degrees from the School of Computer Science and Engineering, Nanjing University of Science and Technology, in 2003 and 2007, respectively. He is now working in the School of Computer Science and Engineering, Nanjing University of Science and Technology as an associate professor. His research interests includes remote-sensing information processing, computer simulation and image processing. (Email: zebin.wu@gmail.com)

WEI Zhihui was born in Jiangsu Province in 1963. He is now working in the School of Computer Science and Engineering, Nanjing University of Science and Technology as a professor. His research interests incudes image/signal processing and wavelet analysis.