An Inner-product Based Discriminative IRLS Algorithm for Sparse Hyperspectral Detection

Zhongwei Huang, Zhenwei Shi, *Member, IEEE* and Yun Zhang, *Member, IEEE*

Abstract

Automatic target detection is an important application in the hyperspectral image processing field. It is currently known that any test pixels in hyperspectral images can be represented within a spectral dictionary using appropriate sparse coefficients. Based on this assumption, some sparsity-based algorithms are developed for hyperspectral detection. This kind of sparse learning method attempts to find the sparse representation from a spectral library, i.e., a dictionary data set from which useful information is extracted. Among these algorithms, the iteratively reweighted least squares (IRLS) strategy is believed to be a simple and useful tool for sparse representation. However, when dealing with the hyperspectral data, the dictionary for sparse learning is usually high-dimensional which dramatically increases the scale and complexity of sparse learning. In such cases, most sparsity-based algorithms including the IRLS strategy lost their efficacy. To deal with this situation, we propose a discriminative IRLS algorithm, called inner-product based discriminative IRLS detector (IDIRLSD), which decreases the scale and complexity problem by discriminatively seeking a sub-dictionary that retains the most critical information. Also IDIRLSD applies a convex minimization for approximately solving the sparse recovery problem. A weighted ℓ_1 minimization is relaxed and solved by IRLS strategy. The proposed algorithm applies an inner-product based function for constructing the small-scale weighted ℓ_1 minimization with respect to the sub-dictionary. The solution provided by IDIRLSD is then applied to label the test pixel as target or background. Experimental results from both synthetic and real hyperspectral data demonstrate the improved efficacy of the proposed algorithm.

Index Terms

Hyperspectral target detection, sparsity-based algorithm, sub-dictionary construction, weighted ℓ_1 minimization, iteratively reweighted least squares.

The work was supported by the National Natural Science Foundation of China under the Grants 61273245 and 91120301, the 973 Program under the Grant 2010CB327904, the Program for New Century Excellent Talents in University of Ministry of Education of China under the Grant NCET-11-0775 and the funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under the Grant VR-2014-ZZ-02.

Zhenwei Shi (Corresponding Author) is with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, PR China, with State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, PR China and also with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, PR China (e-mail: shizhenwei@buaa.edu.cn). Zhongwei Huang was with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, P.R. China. Now, he is with the Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, NB E3B 5A33, Canada (E-mail: zhongwei.huang@unb.ca). Yun Zhang is with the Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, NB E3B 5A33, Canada (E-mail: yunzhang@unb.ca).

I. INTRODUCTION

The hyperspectral imaging (HSI) sensors produce a three-dimensional (3D) data structure called data cube with two spatial dimensions and one spectral dimension [1],[2]. The two-dimensional spatial images are obtained in hundreds of narrow spectral bands by the sensor. The spectrum of each HSI pixel can be viewed as a vector with each of its elements represents the radiance or reflectance value at each spectral band [2],[3]. As each material is characterized by a unique deterministic spectrum, the spectrum of pixel serves as a distinguished feature for discrimination [3]. Many distance measures can be applied to decide whether a given pixel belongs to a kind of material. Because of the high spectral resolution, HSI is more suitable for full pixel detection than the multispectral image [2],[4]. Based on the spectral characteristics of HSI, the target detection for hyperspectral images is a binary classification problem [5]. The spectral information is then used by detection algorithms to label the test pixels as target or background [7].

Several detection algorithms have been developed. Most of them are based on statistical methods. Spectral matched filter (SMF) [8] and constrained energy minimization (CEM) [9],[10] are two well known and widely used algorithms. SMF uses the covariance matrix of the data and CEM uses the similar matrix, so both of them are based on the second-order statistics. However, these algorithms are not stable to deal with complex scenes in which targets of interest are irregularly distributed. In addition, they are sensitive to noise [10].

Recently, sparse signal representation, or sparse recovery [11],[24] has proven to be an extremely powerful tool in many areas including HSI target detection [5],[6]. This success is mainly due to the fact that most natural signals have sparse representations with respect to fixed bases or dictionaries [12]. As this kind of method emphasizes a limited number of spectra in constructing a test pixel, it is usually known as sparse representation or recovery. The sparsity-based algorithms use the spectrum of each test pixel instead of the whole scene for classification, which has an advance of processing irregular distributed target over second-order statistics based algorithms. Sparsity-based method is currently applied in target detection with greedy algorithms (GAs) [11], like orthogonal matching pursuit (OMP) [16] and subspace pursuit (SP) [13]. These greedy algorithms search exhaustively in each iteration to find a sub-optimal approximation, by picking the vector that best correlates with the present residual. But these algorithms may not get the ideal performance since the local search can add to a loss of accuracy.

In this paper, we focus on an alternative method, called convex relaxation technique [14],[15]. This method concentrates on relax the ℓ_0 norm [11] and replaces it by a more tractable approximation [11],[17] such as the ℓ_p norms. For example, in dealing with the relaxed minimization (ℓ_1 or weighted ℓ_1 minimization), the iteratively reweighted least squares (IRLS) strategy [24] applies the weighted squared ℓ_2 norm to adaptively approximate the ℓ_1 norm. This strategy is believed to be an effective algorithm in sparse learning [18]. However, all minimization solvers for HSI processing all face a scale problem which is generated from the high-dimensional information (combined with spectral and spatial information) of HSI sensors provide [19]. Since the IRLS strategy searches for a global optimal solution to the relaxed problem within a huge spectral dictionary, the massive amount of data cast as a challenge for its programming [11]. In order to improve the efficacy of the IRLS strategy for sparse detection

in hyperspectral images, this paper proposes an algorithm that concentrates on discriminatively seeking for an appropriate sub-dictionary which provides a more effective searching process. The conventional IRLS algorithm is developed in order to generate accurate sparse reconstruction with the sub-dictionary. The algorithm we proposed have several advantages which are briefly summarized below.

1. We expound and prove the feasibility of constructing a low-dimensional sub-dictionary to replace the high-dimensional spectral dictionary with respect to given test pixels. A nonlinear inner-product function is proposed for discrimination within the dictionary which generates a sub-dictionary. The discrimination process aims to seek for least but most relevant atoms in the spectra dictionary, and finally uses these atoms to construct a smaller training sample. This approach contributes in cutting the searching effort and calculation when solving the sparse representation problem.

2. Most currently applied sparsity-based algorithms for hyperspectral image fall in the greedy algorithms. The proposed algorithm considers an alternative way for seeking for the sparse representation, called convex relaxation. The algorithm of iteratively reweighted least squares is applied to solve the relaxed problem. The experimental results demonstrate that the proposed algorithm outperforms the greedy algorithms for hyperspectral target detection.

3. It has been assumed that weighted ℓ_1 algorithm outperforms unweighted ℓ_1 algorithm in sparse recovery problem once the weights are properly set [17]. But seldom work is concerned with the weights selection of weighted ℓ_1 algorithm. The proposed algorithm applies IRLS to solve the relaxed weighted ℓ_1 problem, denoted as weighted ℓ_1 IRLS. Performance with respect to the weight selection of proposed algorithm gives support to the efficacy of weighted ℓ_1 algorithm.

The rest of this paper is organized as follows. In Section II, we briefly introduce the HSI target detection technique based on sparse representation. In Section III, the inner-product based discriminative IRLS detector (IDIRLSD) is proposed. In Section IV, Some experimental results are given and Section V draws some conclusions.

II. A Brief Introduction of HSI Target Detection based on Sparse Representation

The spectral dictionary is set as a large matrix with its each column is an M-dimensional vector representing the prior knowledge of spectral signature of one particular material, with M being the number of spectral bands. If N kinds of spectral signatures are included as columns in it, the spectral dictionary $\mathbf{A}$ is a $M \times N$ matrix [5],[20]. Usually in practice we get $M \ll N$. Denote $\mathbf{x}$ as a hyperspectral pixel observation, whose spectral signature is available in the spectral library $\mathbf{A}$. With respect to the location of spectral signature of $\mathbf{x}$, the matrix $\mathbf{A}$ can be separated as $\mathbf{A} = [\mathbf{A}_b \ \mathbf{A}_t]$, where $\mathbf{A}_b$ presents spectral signatures of those background materials and $\mathbf{A}_t$ presents the spectral signature of the target material.

Recent studies have shown that it is possible to reconstruct a test pixel $\mathbf{x}$ with highly incomplete sets of linear measurements within an appropriate dictionary $\mathbf{A}$ [11],[17],[20],[22]. In the sparse recovery problem, the sparse coefficients $\mathbf{s}$ satisfies $\mathbf{As} = \mathbf{x}$. Also the sparsity of vector $\mathbf{s}$ should be minimized, which is defined as the number of nonzero entries in $\mathbf{s}$, i.e. the ℓ_0 norm of vector $\mathbf{s}$:

$$\|\mathbf{s}\|_0 = \#\{i : s_i \neq 0\}. \tag{1}$$

Basically, the representation satisfies $\mathbf{s}$ from $\mathbf{As} = \mathbf{x}$ can be obtained by solving the following sparse reconstruction problem

$$(\mathbf{P}_0) : \widehat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad subject\ to\ \mathbf{As} = \mathbf{x}. \tag{2}$$

When we find the sparse solution $\mathbf{s}$ to the proposed problem, the class [5] of $\mathbf{x}$ can be determined by

$$D(\mathbf{x}) = \|\mathbf{x} - \mathbf{A}_b\mathbf{s}_b\|_2 - \|\mathbf{x} - \mathbf{A}_t\mathbf{s}_t\|_2, \tag{3}$$

where $\mathbf{s}_b$ and $\mathbf{s}_t$ respectively represent the recovered sparse coefficients corresponding to the background sub-dictionary $\mathbf{A}_b$ and target sub-dictionary $\mathbf{A}_t$. If $D(\mathbf{x}) > \delta$, where δ is a given threshold, then $\mathbf{x}$ is labeled as a target pixel; otherwise, it is labeled as a background pixel.

III. Proposed Algorithm

Our method is basically twofold. Firstly, in order to tackle the difficulty of solving the basic $\mathbf{P}_0$ problem, we apply a relaxed model. Secondly, in order to reduce the undesired computational cost brought by the dimensionality of dictionary matrix, we provide a method to construct a sub-dictionary in each iteration of our algorithm, which is suitably designed by improving from the conventional IRLS algorithm.

To start with, the sparsity-based algorithms for target detection necessarily consider the sparse representation model:

$$(\mathbf{P}_0) : \widehat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad subject\ to\ \mathbf{As} = \mathbf{x}. \tag{4}$$

However, a tough issue arises as seeking the optimal solution of $(\mathbf{P}_0)$ is NP-hard. Recent methods all give approximate solvers [11]. A natural way of regularize it is replace ℓ_0 minimization with its best convex approximant—ℓ_1 minimization [22]:

$$(\mathbf{P}_1) : \widehat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad subject\ to\ \mathbf{As} = \mathbf{x}, \tag{5}$$

or weighted ℓ_1 minimization [11],[17]:

$$(\mathbf{WP}_1) : \widehat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{Ws}\|_1 \quad subject\ to\ \mathbf{As} = \mathbf{x}. \tag{6}$$

$\mathbf{W}$ is a diagonal matrix with $w_1, ..., w_N$ as weights on the diagonal and zeros elsewhere.

Their minimization problem, known as $(\mathbf{P}_1)$ and $(\mathbf{WP}_1)$, both are convex problems, which can be solved efficiently using modern techniques [11]. But it has been proposed that the two ℓ_1 relaxations $(\mathbf{P}_1)$ and $(\mathbf{WP}_1)$ will obtain different solutions in general, which suggests that the weights of $(\mathbf{WP}_1)$ may have an important influence on the solution that $(\mathbf{WP}_1)$ obtain. If their values can be set properly, the sparse recovery of convex relaxation can be improved [17]. Thus, the proposed algorithm aims to construct the proper weights base on this assumption to enhance the sparse representation.

In addition, in order to solve the weighted ℓ_1 minimization, we apply the weighted IRLS strategy, which makes a relaxation for the construction $\mathbf{As} = \mathbf{x}$ with ℓ_2 norm and uses Lagrange parameter to covert weighted ℓ_1 minimization into an unconstraint version:

$$(\mathbf{WP}_1^{\lambda}) : \widehat{\mathbf{s}} = \arg\min_{\mathbf{s}} \ \lambda\|\mathbf{Ws}\|_1 + \frac{1}{2}\|\mathbf{x} - \mathbf{As}\|_2^2. \tag{7}$$

In the IRLS algorithm, set $\mathbf{S} = diag(|s_i|)$ and we get:

$$\|\mathbf{s}\|_1 = \mathbf{s}^T\mathbf{S}^{-1}\mathbf{s}. \tag{8}$$

Thus the ℓ_1 norm is approximately computed as a weighted version of the squared ℓ_2 norm [11]. Given a current approximate solution $\mathbf{s}_{k-1}$, IRLS attempts to solve following minimization:

$$(\mathbf{M}_k) : \widehat{\mathbf{s}} = \arg\min_{\mathbf{s}} \ \lambda\mathbf{s}^T\mathbf{W}\mathbf{S}_{k-1}^{-1}\mathbf{s} + \frac{1}{2}\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2. \tag{9}$$

The solution of this problem can be obtained using standard linear algebra. However, the above-discussed spectral dictionary $\mathbf{A}$ is still a high-dimensional data set, which makes the solving of $\mathbf{WP}_1^\lambda$ and $\mathbf{M}_k$ less of efficiency. Thus, another aim of the proposed algorithm is to construct a low-dimensional sub-dictionary to make the problem become small-scale.

A. Analysis of Sub-dictionary Construction

The reason for constructing a sub-dictionary for sparse reconstruction is twofold. Firstly, the spectral library we used as dictionary can be represented by a matrix with each column residing a single spectrum that representing a pure material. The matrix usually contains considerable number of spectra which necessarily cause the number of columns larger than the number of rows within the matrix. Mathematically, this will cause similar and redundancy in the incident matrix and therefore make the sparse reconstruction process harder and time-consuming. Secondly, it is proved the difficulty of general sparse reconstruction task and the main reason for this difficulty goes to the effort needed for searching the "useful" columns that will be used for constructing the given pixel. Thus a direct strategy to simplify this process can be narrowing down the potential columns, in other words, using a sub-dictionary that contains less spectra that are more likely useful for construction of given spectrum. Meanwhile, a straightforward measure on the size of the sub-dictionary can be full-rank, which mathematically enough for construction of any vectors. As a result, the reconstruction within the sub-dictionary is no longer a sparse reconstruction problem, which simplifies the original problem.

In practice, the proposed algorithm applies a nonlinear inner-product based function for discriminatively seeking for a sub-dictionary. Suppose the sub-dictionary is denoted as $\mathbf{A}_{\Lambda_0} \in \mathrm{R}^{M\times K_0}$ and Λ_0 is the index set of K_0 columns of the dictionary $\mathbf{A}$ that are selected to construct the sub-dictionary. Then the sparse recovery coefficients within the sub-dictionary is solved by solving a sub problem. The recovery coefficients for the locations corresponding to the unselected columns of $\mathbf{A}$ will be set as zero. Denote the smaller scale problem $\mathbf{WP}_1^\lambda$ as a sub problem: $\mathbf{WP}_1^\lambda$, then the proposed low-dimensional problem becomes:

$$(\mathbf{SWP}_1^\lambda) : \widehat{\mathbf{s}}_{\Lambda_0} = \arg\min_{\mathbf{s}_{\Lambda_0}} \ \lambda\|\mathbf{W}\mathbf{s}_{\Lambda_0}\|_1 + \frac{1}{2}\|\mathbf{x} - \mathbf{A}_{\Lambda_0}\mathbf{s}_{\Lambda_0}\|_2^2. \tag{10}$$

$\mathbf{W}$ is a diagonal matrix with w_j, $j = 1, ..., K_0$ as weights on the diagonal and zeros elsewhere. $\mathbf{s}_{\Lambda_0}$ is the solution to the sub $\mathbf{WP}_1^\lambda$ problem, a K_0-dimensional vector.

In order to construct and solve the problem we proposed, we first expound and prove the feasibility of sub-dictionary construction. Then the proposed inner-product based function is applied to make discriminative construction of the sub-dictionary. At last, we develop the famous IRLS (iteratively reweighted least squares) strategy to solve the sparse recovery problem on a smaller scale.

The dictionary of training samples $\mathbf{A}$ as mentioned before, can be extremely large which usually contains hundred of spectra. Each spectrum is taken as a column of matrix $\mathbf{A}$. The hyperspectral images have a fixed range of bands and hundreds of spectral signatures are needed for reconstructing the background and target pixel. Thus the rows and columns in the matrix $\mathbf{A} \in \mathrm{R}^{M \times N}$ can be described as $M \ll N$. If it is possible to construct a sub-dictionary with the columns of matrix $\mathbf{A}$, the selected columns must already eligible for representing the M-dimensional test pixel. For this consideration, we attempt to find the upper-bound of the sparsity of the sparest coefficients. Since once this upper-bound is found, we can approximate the number of columns that should be used in the representation, which will provide a measure of the size of our sub-dictionary.

Suppose the $rank$ of matrix $\mathbf{A}$ equals with n, then we rewrite $\mathbf{A}$ as:

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 ... \mathbf{a}_N], \tag{11}$$

where $\mathbf{a}_i$ is the ith column of $\mathbf{A}$, $i = 1, ..., N$. Since $rank(\mathbf{A}) = n$, there are at most n columns that are linear independent. By the definition given by the linear algebra, the vectors:

$$\mathbf{a}_1 \ \mathbf{a}_2 ... \mathbf{a}_N$$

constitute a n-dimensional linear space $\mathbf{V}$.

Now we turn to consider the sparsity of the sparsest solution $\mathbf{s}$. If this sparsity has an upper bound, we can get a prior knowledge of the size of the sub-dictionary we need. Since the sparser the solution is, the less columns of $\mathbf{A}$ are used in reconstruction of $\mathbf{x}$. The $spark$ of a given matrix $\mathbf{A}$ has been used to provide criterion for uniqueness of sparse solutions, i.e., once a solution $\mathbf{s}$ meets this criterion it is the only possible sparest solution to the construction problem $\mathbf{As} = \mathbf{x}$. In algebra, the $spark$ of a given matrix $\mathbf{A}$ is the smallest number of columns from $\mathbf{A}$ that are linearly dependent [11]. Existed researches have been conducted to find the aforementioned criterion using the $spark$ of matrix and proposed that the sparsest possible solution $\mathbf{s}$ to $\mathbf{As} = \mathbf{x}$ must be less than $spark(\mathbf{A})/2$. For detailed proof of this criterion readers are directed to [11]. Mathematically, this is a sound and favorable criterion since it provides the supremum (the least upper bound). However, the computational cost brought by obtaining the $spark$ is undesired compared to the $rank$, as it calls for a combinatorial search over all possible subsets of columns from the matrix. In addition, it will become especially hard to implement sparse learning with hyperspectral data which necessarily calls for a dictionary matrix with high dimensionality.

In practice, it is more favorable to identify a criterion that comes with a general upper bound that easier to obtain. Our idea of finding this upper bound rooted in basic linear algebra theory. For the linear equation $\mathbf{As} = \mathbf{x}$, the function of coefficients $\mathbf{s}$ can be simply described as identify the usage of each column entry within the dictionary matrix $\mathbf{A}$. So it is natural to come up with a set of columns that constitutes a basis which is capable of representing

any given vector **x** by linear combination. Therefore one can tell the least numbers of columns for reconstruction a vector should be less than the number of columns in a basis, which in other words, the $rank$ of the matrix. Combine this criterion with aforementioned supremum criterion, we obtain the following result.

Theorem 1: *In the construction* $\boldsymbol{As} = \boldsymbol{x}$*, the sparsity of the sparsest recovery solution* **s** *must be smaller than* $(\boldsymbol{A})/2$ *and must be no larger than* $rank(\boldsymbol{A})$.

Proof:

Since the supremum has been proved in [11], we give the proof for another upper bound of $rank(\mathbf{A})$.

The recovery formation using $\mathbf{A}$ and $\mathbf{s}$ can be rewritten as:

$$\mathbf{As} = [\mathbf{a}_1\ \mathbf{a}_2 ... \mathbf{a}_N]\ [s_1\ s_2 ... s_N]^T = \mathbf{x}, \tag{12}$$

as $\mathbf{x} \in \mathbf{V}$. Since $rank(\mathbf{A}) = n$, assume the sparsity of the sparsest recovery solution $\mathbf{s}$ equals to k, which satisfies $k > n$, the sparse recovery can be presented as:

$$[\tilde{\mathbf{a}}_1\ \tilde{\mathbf{a}}_2\ ...\ \tilde{\mathbf{a}}_k][\tilde{s}_1\ \tilde{s}_2\ ...\ \tilde{s}_k]^T = \mathbf{x}. \tag{13}$$

Meanwhile, there are at most n columns of $\mathbf{A}$, i.e., a basis of the linear space $\mathbf{V}$, that are linear independent, marked as:

$$[\hat{\mathbf{a}}_1\ \hat{\mathbf{a}}_2 ... \hat{\mathbf{a}}_n]. \tag{14}$$

Any vector within the space $\mathbf{V}$ can be linearly represented by the basis, the columns used in sparse recovery can be presented as:

$$\tilde{\mathbf{a}}_i = [\hat{\mathbf{a}}_1\ \hat{\mathbf{a}}_2 ... \hat{\mathbf{a}}_n][\hat{s}_{1_i}\ \hat{s}_{2_i} ... \hat{s}_{n_i}]^T, \tag{15}$$

$i = 1, ..., k$. Then the sparse recovery can be rewritten as:

$$\Sigma_{i=1}^{k}[\hat{\mathbf{a}}_1\ \hat{\mathbf{a}}_2 ... \hat{\mathbf{a}}_n][\hat{s}_{1_i}\ \hat{s}_{2_i} ... \hat{s}_{n_i}]^T \tilde{s}_i = \Sigma_{j=1}^{n} \hat{\mathbf{a}}_j \bar{s}_j, \tag{16}$$

where $\bar{s}_j = \tilde{s}_1 \hat{s}_{j_1} + \tilde{s}_2 \hat{s}_{j_2} + ... + \tilde{s}_k \hat{s}_{j_k}$, $j = 1, ..., n$. Thus we get a sparser recovery with sparsity equals with $n < k$. So the sparsity of the sparest recovery solution should not be k, the proof is completed.

Among the numbers of solutions to construction problem $\mathbf{As} = \mathbf{x}$, we may verify whether the solution we get is sparse enough by applying this $rank$ criteria which has been applied in practice in proposed algorithm. To be explicit, we can apply this criteria $\|\mathbf{s}\|_0 \leq rank(\mathbf{A})$ as a constrain for the numbers of columns we used and thus provide a desired size of the sub-dictionary. In next part, we will introduce our method to generate the sub-dictionary and a sub $\mathbf{WP}_1^{\lambda}$ minimization problem via our criterion.

B. Construct the Sub $\boldsymbol{WP}_1^{\lambda}$ via Inner-product based Function

In this part we introduce our method to generate the sub-dictionary and the construction of the aforementioned sub $\mathbf{WP}_1^{\lambda}$ problem. In the proposed algorithm, two aspects of the construction of the sub $\mathbf{WP}_1^{\lambda}$ problem should be

concerned— the construction of the sub-dictionary $\mathbf{A}_{\Lambda_0}$ and the construction of weight matrix $\mathbf{W}$. The proposed algorithm solve this two problems using an inner-product based nonlinear function:

$$F(\mathbf{x}, \mathbf{y}) = 1 - \arccos \frac{< \mathbf{x}, \mathbf{y} >}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}. \tag{17}$$

In a broad sense, the function is a distance measure of vector $\mathbf{x}$ and $\mathbf{y}$. In a geometric sense, it describes the relation of vectors by their intersection angles. Once $\mathbf{x}$ and $\mathbf{y}$ are normalized, the function can be rewritten as

$$F(\mathbf{x}, \mathbf{y}) = 1 - \arccos < \mathbf{x}, \mathbf{y} > . \tag{18}$$

The inner-product based function is based on the idea of measure a similar between two high-dimensional vectors. The larger the function value F is, the larger the similar between two vectors is.

To deal with the sparse recovery problem (2), each column of dictionary $\mathbf{A}$ in construction can be discriminated by the measurement $F(\mathbf{a}_i, \mathbf{x})$. To be explicit, once the value of the function is large enough, the column $\mathbf{a}_i$ is believed to be eligible as a candidate for the sub-dictionary.

Consider the construction of weight matrix for the sub $\mathbf{WP}_1^\lambda$ problem. We developed the idea introduced by Candes et al. [17], who proposed a solver for reweighted ℓ_1 minimization which update the weights in each iteration using the reciprocal of the last solution. The reweighted ℓ_1 minimization is proved to be progressive for enhancing the sparse representation. As different weights allow for different penalize for each location, if one location of the coefficients, marked as s_i, is known to be a considerable large nonzero entry, then the corresponding weight for this location should be smaller.

Note that since we aim to develop IRLS [11] to solve the weighted ℓ_1 minimization, our model is different than that of [17]. But we incorporate their idea into the construction of weights in our weighted ℓ_1 model. When solving the sub $\mathbf{WP}_1^\lambda$ problem, the weights for each location is not uniformly distributed in the proposed algorithm. As the inner-product function we proposed provides a measure of the weights of different recovery coefficients in the sparse representation, the function value is applied for constructing the weight matrix for the sub $\mathbf{WP}_1^\lambda$ problem. The above-described Λ_0 is still the index set of K_0 columns of the dictionary $\mathbf{A}$ used for construction of the sub-dictionary. Let $\Lambda_0(j)$ be its jth member. The function value $F(\mathbf{a}_i, \mathbf{x}) = 1 - \arccos < \mathbf{a}_i, \mathbf{x} >$ is marked as σ_i, $i = 1, ...N$. Then the function value of $F(\mathbf{a}_{\Lambda_0(j)}, \mathbf{x})$ is marked as $\sigma_{\Lambda_0(j)}$, $j = 1, ...K_0$. We compute the weights on the diagonal of $\mathbf{W}$ successively as:

$$w_j = 1/(\sigma_{\Lambda_0(j)} + \varepsilon), \tag{19}$$

$j = 1, ..., K_0$, where ε is a parameter whose value provides different degrees of penalization.

C. Inner-product Based Discriminative IRLS Detector

Based on the analysis above, the proposed algorithm inner-product based discriminative IRLS detector (IDIRLSD) is mainly consists of two parts: discriminatively constructing the sub $\mathbf{WP}_1^\lambda$ problem with inner-product based function and solving the sub $\mathbf{WP}_1^\lambda$ problem using IRLS strategy. Thus the inner-product based discriminative IRLS detector (IDIRLSD) is obtained as follows:

Algorithm Outline: IDIRLSD

(1) Normalize each column of dictionary $\mathbf{A}$: $\mathbf{a}_1,\ \mathbf{a}_2...\mathbf{a}_N$ and the given pixel $\mathbf{x}$. $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2...\mathbf{a}_N]$ as we discussed above.

(2) Compute the function: $\sigma_i = F(\mathbf{a}_i, \mathbf{x}) = 1 - \arccos < \mathbf{a}_i, \mathbf{x} >,\ i = 1, ...N$ successively. Then descending sort the sequence: $\Delta = [\sigma_1...\sigma_N]$, and get the index set Λ_0=$\{K_0$ indices corresponding to the K_0 largest numbers in $\Delta\}$. Let $\Lambda_0(j)$ $(j = 1, ..., K_0)$ denotes the jth index in Λ_0.

(3) The sub-dictionary is then constructed as $\mathbf{A}_{\Lambda_0}$, and the corresponding low-dimensional weights matrix $\mathbf{W}$ is set as $w_j = 1/(\sigma_{\Lambda_0(j)} + \varepsilon)$. $(j = 1, ..., K_0)$.

(4) Solve the relaxed smaller-scale minimization

$$(\mathbf{SWP}_1^{\lambda}) : \widehat{\mathbf{s}}_{\Lambda_0} = \arg\min_{\mathbf{s}_{\Lambda_0}} \ \lambda\|\mathbf{W}\mathbf{s}_{\Lambda_0}\|_1 + \frac{1}{2}\|\mathbf{x} - \mathbf{A}_{\Lambda_0}\mathbf{s}_{\Lambda_0}\|_2^2.$$

Initialization: Initialize k=0, and set

the initial approximation $\mathbf{s}_0$ = **1**, where **1** stands for a vector with all its entries being 1.

the initial iterative weight matrix $\mathbf{S}_0$ = **I**, where **I** stands for an identity matrix.

Main Iteration: Increment k by 1, and apply these steps:

Regularized Least Squares: Solve the linear system

$$(2\lambda\mathbf{W}\mathbf{S}_{k-1}^{-1} + \mathbf{A}_{\Lambda_0}^T\mathbf{A}_{\Lambda_0})\mathbf{s}_{\Lambda_0} = \mathbf{A}_{\Lambda_0}^T\mathbf{x}$$

iteratively, producing result $\mathbf{s}_{\Lambda_0}^k$.

Iterative Weight Matrix Update: Update the diagonal weight matrix $\mathbf{S}$ using $\mathbf{s}_{\Lambda_0}^k$: $S_k(j,j) = |\mathbf{s}_{\Lambda_0}^k(j)| + \xi$.

Stopping Rule: If $\|\mathbf{s}_{\Lambda_0}^k - \mathbf{s}_{\Lambda_0}^{k-1}\|_2$ is smaller than some predetermined threshold, stop. Otherwise, apply another iteration.

Output: The desired result is $\mathbf{s}_{\Lambda_0}^k$, marked as $\widehat{\mathbf{s}}_{\Lambda_0}$.

Let $\widehat{\mathbf{s}}_{\Lambda_0}(j)$ denotes the jth element of $\widehat{\mathbf{s}}_{\Lambda_0}$, then the sparse recovery coefficients $\mathbf{s} = \{s_i\}_{i=1}^N$ is obtained as:

$$s_i = \begin{cases} \widehat{\mathbf{s}}_{\Lambda_0}(j) & ,\ \exists\ j = 1, ..., K_0,\ i = \Lambda_0(j), \\ 0 & ,\ else. \end{cases}$$

(5) Rewrite the sparse coefficients $\mathbf{s}$ as [$\mathbf{s}_b\ \mathbf{s}_t$], corresponding with locations of target and background sub-dictionaries. Apply (3) for determining whether the given pixel is a target pixel.

Note that the proposed IDIRLSD is a derived version of IRLS algorithm. Usually, IRLS algorithms solve a weighted ℓ_1 minimization using a reweighted strategy [11]. In the case of IDIRLSD described above, The reweight matrix $\mathbf{S}$ is updated within each iteration while the outside weight matrix $\mathbf{W}$ for the weighted l1 minimization is fixed as described in [11].

The number of iterations IDIRLSD needed is equals to the image pixel numbers, which is the same with the conventional IRLS [18],[24] or ℓ_1 minimization [11]. However, the improvement of IDIRLSD, compared with these conventional convex relaxation based algorithms, is twofold: it reduces the scale of the minimization problem within each iteration and it applies a novel way for constructing the weighted ℓ_1 minimization model which will enhance the accuracy of sparse recovery. The efficacy of the proposed IDIRLSD algorithm will be demonstrated with several experiments based on hyperpsectral data in next section.

IV. Experimental Results

In this section, we use two synthetic hyperspectral images and two real hyperspectral images for target detection experiments. The proposed algorithm IDIRSLD is compared with six algorithms including two statistical methods: spectral matched filter(SMF) [8], constrained energy minimization (CEM) [9],[10] and four sparsity-based algorithms: orthogonal matching pursuit(OMP) [16], subspace pursuit (SP) [13], unweighted ℓ_1 minimization [11] and conventional IRLS [11]. OMP and SP fall in the general class of greedy algorithms [11] while unweighted ℓ_1 minimization and IRLS are compared as classical convex relaxation methods [11],[24]. For those two statistics based algorithms, the CEM is implemented following the description in [10]. In the implementation of SMF, we used the regularization method introduced by [8], which adjusted background covariance matrix to make its inverse more stable.

In order to use sparsity-based algorithms to conduct synthetic and real data experiments, the first task is to obtain the dictionary matrix [5]. In hyperspectral target detection, the spectral dictionary $\mathbf{A}$ is constructed with two parts: the background sub-dictionary $\mathbf{A}_b$ and the target sub-dictionary $\mathbf{A}_t$. For the background library, we used the United States Geological Survey (USGS) digital library [23]. The reflectance value of 498 materials for 189 spectral bands distributed in the interval 0.4-2.5 μm. After adding the spectrum of the target pixel, which is considered as a prior knowledge used as $\mathbf{A}_t$, the whole spectral library is $\mathbf{A}^{189\times 499}$. For preparation, each column of the spectral dictionary is normalized as well as the test sample pixels. The main reason for this preparation is for the purpose of avoiding the scale difference that may exist between the dictionary and the test spectral pixels.

To apply the proposed algorithm, inner-product based discriminative IRLS detector (IDIRLSD), several parameters are particularly concerned. The sparsity parameter K_0 is set to $rank(\mathbf{A})/\alpha$, where α is first set to 2, then rises within a range that guarantees satisfactory result. In our experiments, different values of α within the range of 2 to 10 get great performance. Also, the weight parameter ξ for IRLS iteration is a small quantities within the range of 0 to 1. The Lagrange parameter λ is set to 0.095, based on the observation of Elad et al [24]. In order to demonstrate the advanced efficacy of weighted ℓ_1 IRLS over the unweighted one, we compared the performances of the different set of parameter ε for $w_i = 1/(\sigma_{\Lambda_0(j)} + \varepsilon)$ in the real image experiment 2. Since the ℓ_1 minimization [22] is believed to be a useful tool in solving the sparse recovery problem, we also include it for comparison. The problem (P_1) was cast as a linear programming (LP) problem in the experiments and solved using modern interior-point methods as described in [11]. Two other sparsity-based greedy algorithms OMP [16] and SP [13] are also implemented for comparison. Note that the sparsity parameters within all the mentioned sparsity-based

algorithms are adjusted with respect of each experiment data to get the best performance of its own. Usually the Lagrange parameter λ is the most importance parameter to adjust the sparsity level of the result within sparsity-based algorithms. However for the proposed IDIRLSD algorithm, a constraint on the size of the sub-matrix is applied to avoid the underdetermined problem. As a result, the reconstruction within the sub-dictionary is no longer a sparse reconstruction problem. In this case the sparsity conducted as $K_0 = rank(\mathbf{A})/\alpha$ will be a determinant factor for the level of sparsity. In practice, the sparsity K_0 has been adjusted together with the Lagrange parameter λ in order to get the best performance. In our observation the best performance can always be achieved with Lagrange parameter equals to 0.095 in our experiments which agrees with [24]. In addition, in order to suggest the advantage of the proposed algorithm IDIRLSD over the original IRLS algorithm, we also compare these two algorithms in the real image experiment 2. The IRLS algorithm is implemented as reweighted ℓ_1 minimization as described in [11]. In the proposed algorithm IDIRLSD and other sparsity-based algorithms, once a sparse representation vector is obtained, we apply (3) to compare the recovery level of using $\mathbf{A}_b$ and $\mathbf{A}_t$ respectively which lead to a binary classification.

The results of these algorithms are compared both visually and quantitatively by the receiver operating characteristics (ROC) curves [25]. The ROC curve is applied for quantitative analysis of the detection results. ROC curve is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. The true pixel information is provided by ground truth (the real distribution of target and background). As the threshold varying in the whole possible region, the ROC curve is generated by plotting PD (probability of detection) as a function of PFA (probability of false alarms). PD is calculated by the ratio of the number of target pixels that are labeled as targets and the total number of true target pixels. PFA is calculated by the ratio of the number of background pixels that are labeled as targets and the total number of pixels. ROC analysis provides tools to compare classifiers quantitatively, i.e., the lager the area embraced by the plot, the better the classifier performs.

A. Synthetic Image Experiment 1

We design synthetic hyperspectral with the synthetic image designed method introduced by Chang et al. [26]. In the synthetic image, five mineral pure pixels are used for target detection, respectively are named as copiapite (Co), actinolite (Ac), barite (Ba), chrysocolla (Ch) and axinite (Ax). The background of the image was simulated by the mean spectrum of an entire real image collected by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor, shown in Fig. 1. The mean spectrum of the entire real image denoted by M was calculated. Fig. 2 shows the normalized spectra of Co, Ac, Ba, Ch, Ax and M. We used these six kinds of spectra to design a 189-band synthetic image with size of 200×200 called synthetic image 1, where target was simulated by spectra of Co, and the background was simulated by M. Fig. 3 shows the first band of the synthetic image 1 and the ground truth of targets in the synthetic image 1. In the synthetic image 1, there are twenty five target panels arranged in a 5×5 matrix. In column 1 and column 2, from top to bottom, there are five 4×4 and 2×2 pure-pixel panels simulated by spectra of Co, Ac, Ba, Ch and Ax, respectively. Five 2×2 mix-pixel panels are located in column 3, and five 1×1 sub-pixel panels are located in column 4 and column 5, respectively. An additive white Gaussian noise was

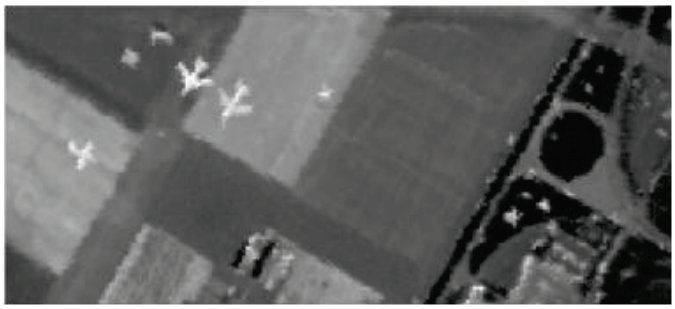


Fig. 1. The first band image of a real AVIRIS hyperspectral image.

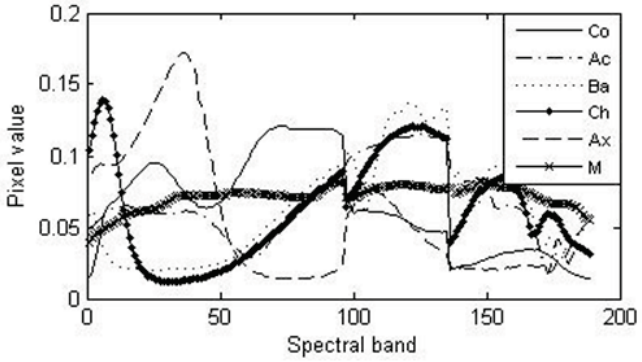
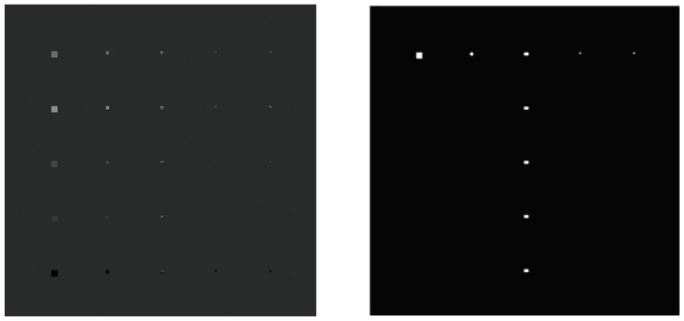


Fig. 2. Spectra of Co, Ac, Ba, Ch, Ax and M.

added to the synthetic image 1 to achieve a 60 dB signal-to-noise ratio (SNR).

The synthetic image 1 was used to detect the mineral target: Co. For the two statistical methods, spectrum of Co as shown in Fig. 2 were used as the prior knowledge of spectral signatures of detected target in the CEM. Fig. 4 shows ROC curves of these algorithms, the performances except that of unweighted ℓ_1 minimization are generally satisfactory. The proposed algorithm IDIRLSD and statistical algorithms (SMF and CEM) get the best results since



(a) (b)

Fig. 3. (a) The first band image of the synthetic image 1. (b) The ground truth of targets of the synthetic image 1.

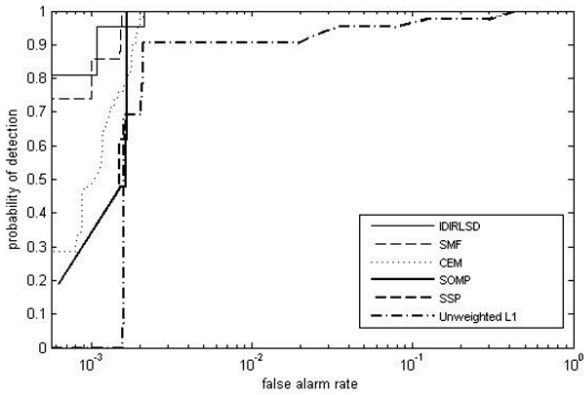
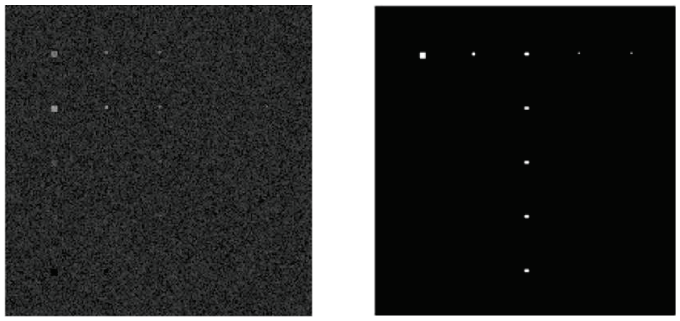


Fig. 4. ROC curves of different algorithms for the synthetic image 1.



(a) (b)

Fig. 5. (a) The first band image of the synthetic image 2. (b) The ground truth of targets of the synthetic image 2.

their ROC curves are always 1 when the false alarm rate just passes 10^{-3}. The ROC curves suggest a strong competitiveness of sparsity-based algorithms against classic statistical algorithms. However, the ℓ_1 minimization is also proved for its loss of accuracy. Finally, the good performances of algorithms are also due to the high SNR, which suggests there is almost no noise.

B. Synthetic Image Experiment 2

The synthetic image 2 is similar to the synthetic image 1. The difference was that a stronger additive white Gaussian noise was added to the synthetic image 2 to achieve a 20 dB SNR. Fig. 5 shows the first band image of the synthetic image 2 and the ground truth of targets of the synthetic image 2. Fig. 6 shows ROC curves of different algorithms. Fig. 6 shows the ROC curves of the IDIRLSD, SMF and CEM are still much higher than the other algorithms and IDIRLSD has the best performance overall. Finally, under strong noise environment the IDIRLSD has a good performance.

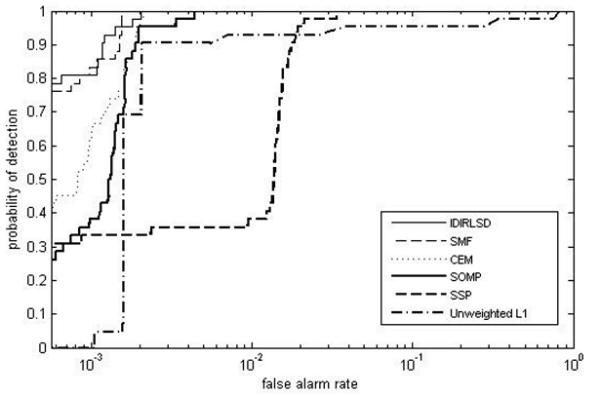


Fig. 6. ROC curves of different algorithms for the synthetic image 2.

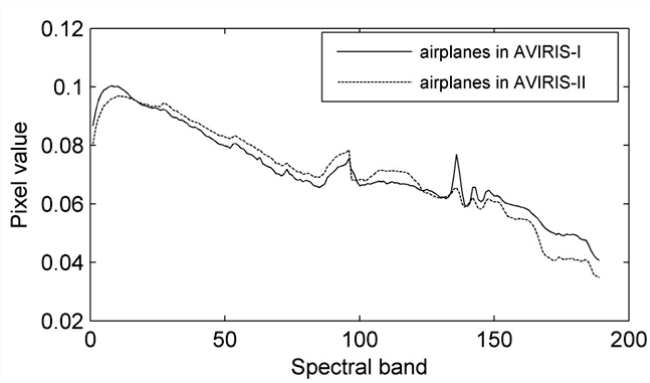


Fig. 7. The prior knowledge of spectral signatures of the two kinds of airplanes in the AVIRIS images.

C. Real Image Experiment 1

In this section, we used the real hyperspectral image data to compare the proposed IDIRLSD algorithm with other algorithms. For real images, every pixel on the targets is considered as a target pixel. The real hyperspectral image was collected by the AVIRIS sensor. The scene was a part of the airport in San Diego, America. The AVIRIS sensor collects the spectral data in 224 bands, and the spectral range is 0.4-2.5 μm. We have removed the water absorption and low SNR bands, and 189 available bands were left. There are two kinds of airplanes in the image. The first kind of three airplanes locate at the upper left corner of the image, and the second kind of three airplanes locate at the lower right corner of the image. Fig. 7 shows the prior knowledge of spectral signatures (normalized) of the two kinds of airplanes. In order to detect the first kind of planes as targets, we used the image named AVIRIS-I. The first band image of the real image AVIRIS-I is shown in Fig. 8 with size of 120$\times$120. In order to compare the different detection results of different algorithms efficiently, we transformed the detection results of different algorithms to binary images with a predefined threshold 0.5, as shown in Fig. 9(a)-(f). From the binary images we can see that IDIRLSD could detect the target more effectively. Almost all of the target pixels are detected.

Fig. 8. The first band image of the real image AVIRIS-I.

The ROC curves of different algorithms for the AVIRIS-I image are shown in Fig. 10. It can be found that in the same false alarm rate, the probability of detection of the IDIRLSD is always higher than those of the other experimental algorithms. Thus, the proposed IDIRLSD performs better than the other experimental algorithms. It is noticeable that other sparsity-based algorithms unweighted ℓ_1 minimization, OMP and SP also suggest a advantage over conventional statistics based CEM and SMF especially in the low false alarm rate region which is important since it suggests the ability of the detection algorithm to find the target with more similar distractions. Meanwhile a step phenomenon can be observed in almost each curve in Fig. 10. which is mainly caused by the relatively less pixel number in the AVIRIS-I.

D. Real Image Experiment 2

In this section, the second kind of planes is detected in the real image AVIRIS-II. The first band of image AVIRIS-II is already shown in Fig. 1 with size of 90×200. The detection within image AVIRIS-II is more difficult since the scene is more complex and contains more background pixels whose signatures are similar with that of the target. The detection results of different algorithms are shown in Fig. 11(a)-(f). The ROC curves of different algorithms are shown in Fig. 12 (the upper-right part has been zoomed in order for better illustration). In this case, Fig. 11 showed that the classical second-order statistical algorithms SMF and CEM performed not so well as they did in synthetic images. Some target pixels were missed and more non-target pixels were detected mistakenly. The sparsity-based algorithms relatively have a better performance, in the Fig. 11, we can see proposed algorithm IDIRLSD detects pixels of target more completely. Also, the ROC curves shown in Fig. 12 denote that when the false alarm rate is smaller than 10^{-3}, the ROC curves of OMP, SP and unweighted ℓ_1 are almost overlapped and the possibility of detection of IDIRLSD is always higher and first curve to reach 1 as false alarm rate increases. Thus, the IDIRLSD has the best detection result on the whole. The result also suggests the advantage of convex relaxation methods over greedy based algorithms. As illustrated, the ROC curves of IDIRLSD and unweighted ℓ_1 minimization are higher than those of OMP and SP especially in low false alarm rate regions. This is mainly because the AVIRIS-II is a relatively large image and with more complex scenes. The greedy algorithms can be

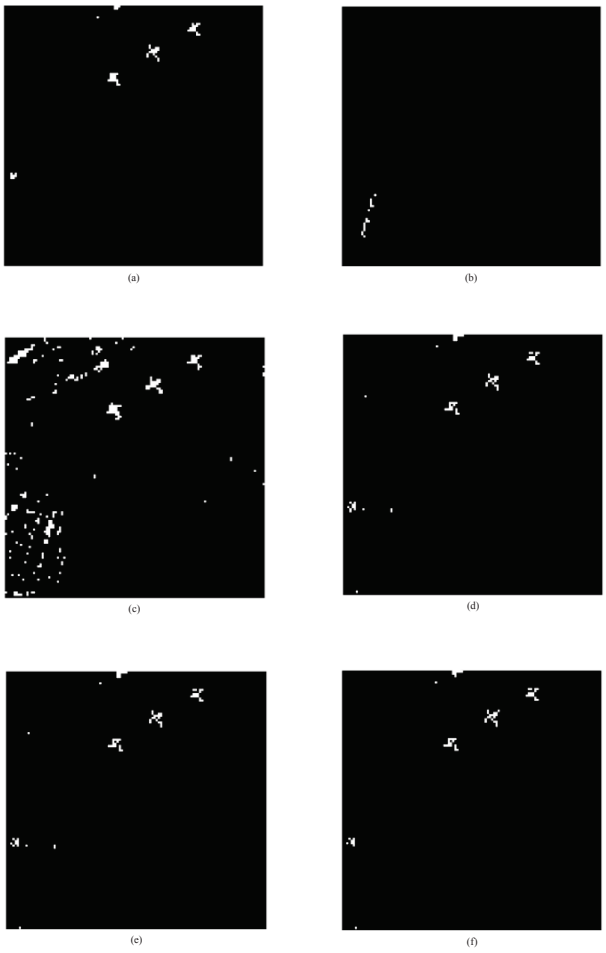


Fig. 9. (a) The detection result of the IDIRLSD for the AVIRIS-I image. (b) The detection result of the SMF for the AVIRIS-I image. (c) The detection result of the CEM for the AVIRIS-I image. (d) The detection result of the OMP for the AVIRIS-I image. (e) The detection result of the SP for the AVIRIS-I image. (f) The detection result of the unweighted ℓ_1 for the AVIRIS-I image.

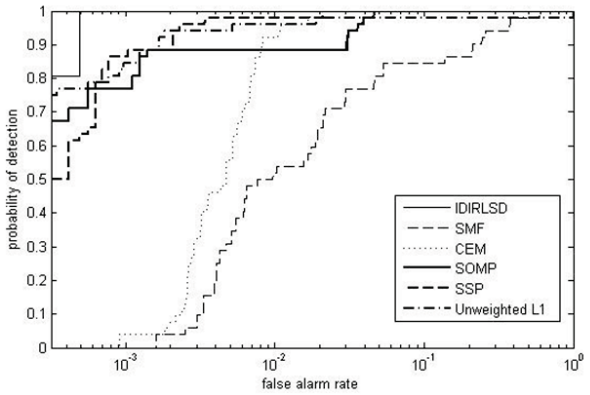


Fig. 10. ROC curves of different algorithms for the AVIRIS-I image.

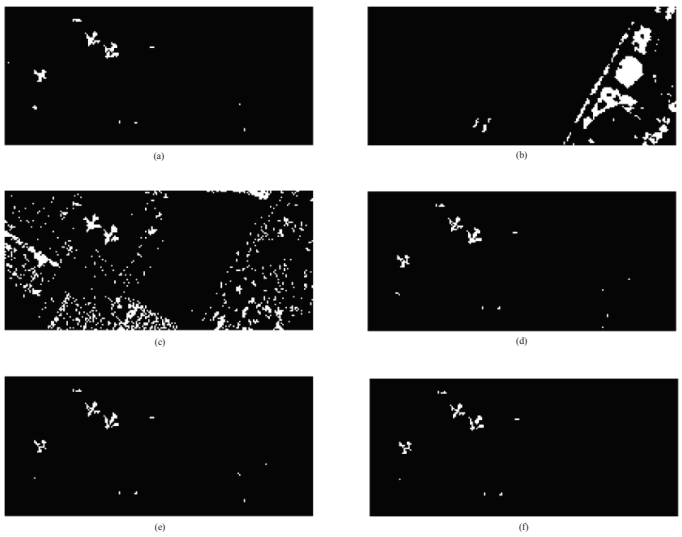


Fig. 11. (a) The detection result of the IDIRLSD for the AVIRIS-II image. (b) The detection result of the SMF for the AVIRIS-II image. (c) The detection result of the CEM for the AVIRIS-II image. (d) The detection result of the OMP for the AVIRIS-II image. (e) The detection result of the SP for the AVIRIS-II image. (f) The detection result of the unweighted ℓ_1 for the AVIRIS-II image.

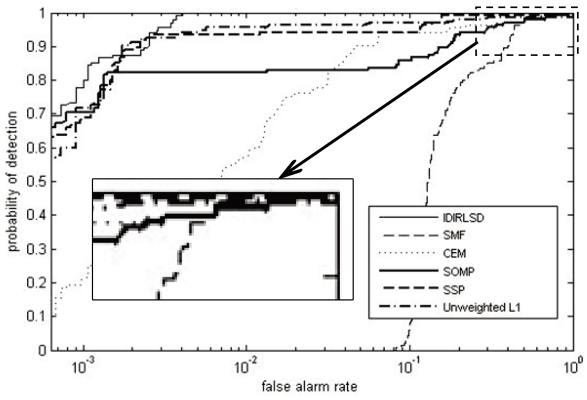


Fig. 12. ROC curves of different algorithms for the AVIRIS-II image.

trapped into local minimum as solutions while convex relaxation based algorithms avoids this problem by directly search for global minimum.

In addition, we include the conventional IRLS algorithm [18],[24] for comparison. The ROC curves for the IRLS and IDIRLSD algorithms are illustrated respectively in Fig. 13. From the ROC curves we can find the ROC of IDIRLSD are always higher. Furthermore, we have compared the time cost when implement these two algorithms IRLS and IDIRLSD on AVIRIS-II. The implementations are under same Matlab environment and the time cost of IRLS is 18.5 times that of IDIRLSD. The IDIRLSD has a better performance which suggests a great advantage over the conventional IRLS algorithm. The reasons for this advantages are twofold: firstly, the discriminative construction

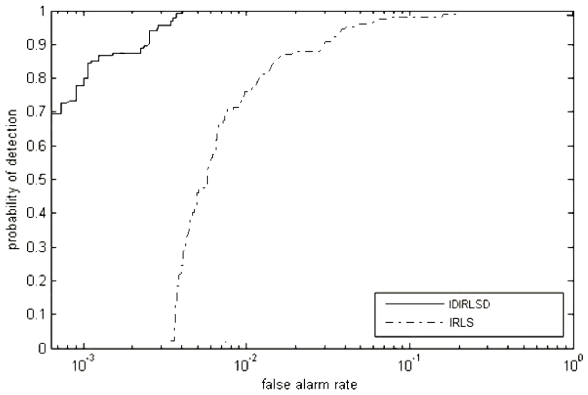


Fig. 13. ROC curves of IRLS and IDIRLSD algorithms for the AVIRIS-II image.

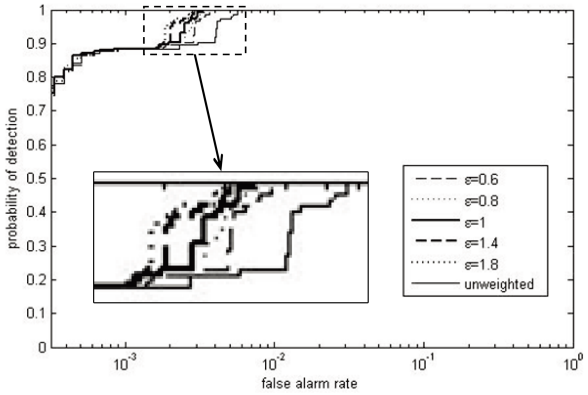


Fig. 14. ROC curves of different set of parameter ε for the AVIRIS-II image.

of dictionary reduces the scale of sparse recovery problem; secondly, the effective construction of the weights for weighted ℓ_1 minimization enhances the accuracy of sparse recovery.

Finally, in order to demonstrate the advanced efficacy of weighted ℓ_1 IRLS over the unweighted one, we compared the performances of the different set of parameter ε for $w_i = 1/(\sigma_{\Lambda_0(j)} + \varepsilon)$ and the unweighted condition on the AVIRIS-II image. The ROC curves with respect to the different sets of parameter ε are shown in Fig. 14 (the upper-left part has been zoomed in order for better illustration). We can see that different curves start by almost overlapped with each other, but the curve with larger ε value within the range 0.6 to 1.8 locates at 1 much faster and the performance is therefore better. Thus, the result suggests that weighted ℓ_1 IRLS has better performance over unweighted one.

E. Experimental Results Analysis

Two synthetic images and two real images were used to do experiments. For the synthetic image 1, due to the fact that the background is very simple and there is nearly no noise in the image, all algorithms perform well. For the synthetic image 2 and the real image AVIRIS-I and AVIRIS-II, the IDIRLSD performs better than other algorithms, although all methods can detect targets well for the synthetic image 2 and real image AVIRIS-I. Since the second-order statistics is suitable to describe Gaussian data, however, in the AVIRIS-II image, target pixels of interest only occupy a few pixels, and the spectra of target pixels do not follow Gaussian distribution. Sparsity-based algorithms do not have this limitation since they concentrated on finding the sparsest recovery coefficients for each test pixel. Once the appropriate sparse solution is obtained, the test pixel can be easily separated with respect to the training dictionary. Also among the sparsity-based algorithms, the proposed algorithm IDIRLSD performs best. This is due to its efficacy on finding the sparse recovery. Finally, the experimental results demonstrate that weighted ℓ_1 IRLS has better performance over unweighted one.

V. Conclusions

Most existing detection algorithms fall in the second-order statistic algorithms, which always face the obstacles in dealing with the irregular distribution of target and additional noise. In this paper, we propose a novel way of hyperspectral detection using sparsity-based algorithm. Currently sparsity-based detection algorithms are mostly greedy algorithms and seldom of them solve the large scale problems effectively. This paper provides a novel way for sparse detection using convex relaxation. Also the proposed algorithm constructs a sub-dictionary to reduce the problem scale of sparse learning. Experiments on both synthetic and real hyperspectral data show that the proposed algorithm performs better than the second-order statistics based algorithms, the sparsity-based greedy algorithms and conventional convex relaxation algorithms for the experimental data in the paper.

References

[1] D. Manolakis, R. Lockwood, T. Cooley, and J. Jacobson, "Is there a best hyperspectral detection algorithm?" in *Proc SPIE*, Orlando, USA, 2009.

[2] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.,* vol. 14, no. 1, pp. 79-116, 2003.

[3] D. Manolakis, and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.,* vol. 19, no. 1, pp. 29-43, 2002.

[4] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 44, no. 6, pp. 1575-1585, 2006.

[5] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Simultaneous joint sparsity model for target detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.,* vol. 8, no. 4, pp. 676-680, 2011.

[6] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral Image Classification Using Dictionary-Based Sparse Representation," *IEEE Trans. Geosci. Remote Sensing,* vol. 49, no. 10, pp. 3973-3985, Oct. 2011

[7] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. New York: Kluwer Academic/Plenum Publishers, 2003.

[8] N. M. Nasrabadi, "Regularized spectral matched filter for target detection in hyperspectral imagery," *IEEE International Conference on Image Processing, ICIP 2007,* pp. IV105-IV108, 2007.

[9] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the coeur d'Alene river valley, Idaho, through the use of a constrained energy minimization technique," *Rem. Sens. Environ.,* vol. 59 no. 1 pp. 64-76, 1997.

[10] H. Ren, Q. Du, C.-I. Chang, and J. O. Jensen, "Comparison between constrained energy minimization based approaches for hyperspectral imagery," *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data,* pp. 244-248, 2003.

[11] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *Society for Industrial and Applied Mathematics,* vol. 51, no. 1, pp. 34-81, 2009.

[12] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE,* vol. 98, no. 6, pp. 1031-1044, 2010.

[13] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory,* vol. 55, no. 5, pp. 2230-2249, 2009.

[14] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[15] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory,* vol. 52, no. 3, pp. 1030-1051, 2006.

[16] J. A. Tropp and A. C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory,* vol. 53, no. 12, pp. 4655C4666, 2007.

[17] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877-905, 2008.

[18] R. Chartrand and Wotao Yin, "Iteratively reweighted algorithms for compressive sensing," *2008 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP,* pp. 3869-3872, Mar. 2008.

[19] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.,* vol. 10, no. 4, pp. 584-597, 2001.

[20] A. Castrodad, Z. Xing, J. Greer, E. Bosch, L. Carin, and G. Sapiro, "Discriminative sparse representations in hyperspectral imagery," *17th IEEE International Conference on Image Processing, ICIP2010,* pp. 1313-1316, Sep. 2010.

[21] O. Egecioglu, H. Ferhatosmanoglu, and U. Ogras, "Dimensionality reduction and similarity computation by inner-product approximations," *IEEE Trans. Knowl. Data Eng.,* vol. 16, no. 6, pp. 714-726, 2004.

[22] T. T. Cai, G. Xu, and J. Zhang, "On recovery of sparse signals via ℓ_1 minimization relax," *IEEE Trans. Inf. Theory,* vol. 55, no. 7, pp. 3388-3397, 2009.

[23] R. N. Clark, G. A. Swayze, A. J. Gallagher, T. V. V. King, and W. M.Calvin, "The U.S. geological survey digital spectral library: Version 1: 0.2 to 3.0 microns," *U.S. Geological Survey, Denver, CO, Open File Rep.,* pp. 93-592, 1993.

[24] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer New York Dordrecht Heidelberg London, 2010.

[25] C.-I. Chang, "Multiparameter receiver operating characteristic analysis for signal detection and classification," *IEEE Sensors J.,* vol. 10, no. 3, pp. 423-442, 2010.

[26] Y.-C. C. Chang, H. Ren, C.-I. Chang, and R. S. Rand, "How to design synthetic images to validate and evaluate hyperspectral imaging algorithms," in *Proc. SPIE Int. Soc. Opt. Eng.,* Orlando, FL, United States, 2008.