A regularized non-Gaussianity based multiple-target detector for hyperspectral images ☆

Zhenwei Shi [a,b,c,*], Shuo Yang [a], Zhen Qin [a]

[a] Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, PR China
[b] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, PR China
[c] Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, PR China

ARTICLE INFO

ABSTRACT

Automatic target detection is an important application in the hyperspectral image processing field. Several target detection algorithms have been developed for hyperspectral images. However, most target detection algorithms were designed to detect one kind of target, and the number of multiple-target detection algorithms is very limited. Besides, the existing multiple-target detection algorithms use second-order statistics, which could characterize Gaussian data well. But for real hyperspectral images, spectra of targets usually do not follow Gaussian distribution. Under such circumstances, we propose a novel multiple-target detection algorithm, named regularized non-Gaussianity based multiple-target detector (RNGMD), which uses the non-Gaussianity statistics to characterize the statistical characteristics of targets' spectra. The RNGMD turns the multiple-target detection into a constrained optimization problem, and utilizes the gradient descent method to solve the optimization problem. Also, we prove the stability of the algorithm. The experimental results demonstrate that the proposed algorithm is more effective than second-order statistics based algorithms.

1. Introduction

The hyperspectral image, with two spatial dimensions and one spectral dimension, is a three-dimensional data called "data cube" [1,2]. The hyperspectral imaging sensor provides each pixel a continuous spectrum with hundreds of very narrow bands, and pixels in the same spectral band form a two dimension image [1,3]. A prominent advantage of the hyperspectral image is the high spectral resolution (the spectral resolution is usually about 10 nm) [4–6]. Because of the high spectral resolution, the hyperspectral image is more suitable for target detection than the multispectral image, and automatic target detection has become an important application in the hyperspectral image processing field.

Several target detection methods have been proposed, but most of them could detect only one kind of target at a time, such as the constrained energy minimization (CEM) detector [7,8], the matched filter (MF) [2,9,10], and the kernel based methods [11–13]. The number of multiple-target detection methods is very limited. Chang et al. [14,15] extended the single-target detection CEM detector to three multiple-target detection approaches, namely multiple-target CEM (MTCEM), sum CEM (SCEM) and winner-take-all CEM (WTACEM). The MTCEM extends the single constraint in CEM to multiple constraints. The SCEM sums several CEM results together, and the WTACEM uses the maximum value of several CEM results. Target-constrained interference-minimized filter (TCIMF) [14] can also be used to detect multiple targets. The TCIMF detects desired target signatures and eliminates the undesired signatures at the same time. Although these methods could detect multiple targets, the detection results are sensitive to noise and are not so practical [15].

The three CEM based multiple-target detection approaches and the TCIMF all use the correlation matrices (second-order statistics) of the data. The second-order statistics could describe Gaussian distribution followed data well. However, in real hyperspectral images, spectra of targets usually do not follow Gaussian distribution [16,17]. In this case, second-order statistics could not capture targets' spectral statistical characteristics well, while some non-Gaussianity statistics could describe the spectral characteristics of targets better. Under such circumstances, we propose a new multiple-target detector which uses non-Gaussianity statistics and

regularization terms to build an objective function and turns the multiple-target detection problem into an optimization problem. It is worthwhile to give the contributions of this paper here:

1. The number of multiple-target detection algorithms is very limited. We propose a novel multiple-target detection algorithm, which could include the MTCEM and the CEM as special cases.

2. The multiple-target detection algorithms available now use the second-order statistics, which suitably describe the Gaussian data. But for real hyperspectral images, spectra of targets usually do not follow Gaussian distribution. In order to deal with this situation, our algorithm utilizes the non-Gaussianity statistics to construct the detector.

3. The proposed algorithm uses the prior knowledge of spectral signatures of interesting targets to construct regularization terms. We only need to change the regularization terms for detecting different kinds of targets.

4. The stability theorem of our algorithm is given. The theorem proves that under certain conditions, the objective function of our algorithm has a certain local minimum.

The rest of this paper is organized as follows. In Section 2, we propose the novel algorithm. The stability analysis of the algorithm is given in Section 3. In Section 4, some experimental results are given, and Section 5 draws some conclusions.

2. Proposed algorithm

The spectrum of each pixel in an N-pixel hyperspectral image can be arranged as a vector as $\mathbf{x}(n) = [x_1^n, \ldots, x_L^n]^T$ ($n = 1, \ldots, N$), where T denotes matrix transposition, and L is the number of spectral bands. Assume that there are m kinds of targets to be detected, and let $\mathbf{d}_i = [d_1^i, \ldots, d_L^i]^T$ ($i = 1, \ldots, m$) represent the prior knowledge of spectral signatures of m kinds of interesting targets. $\mathbf{d}_i$ ($i = 1, \ldots, m$) are available from the spectral library. The spectra of an N-pixel hyperspectral image can be arranged in a matrix form as $\mathbf{X} = [\mathbf{x}(1) \ldots \mathbf{x}(N)]$. To be simple, we assume that the mean of the data has been removed, which means $E\{\mathbf{x}\} = (1/N) \sum_{n=1}^{N} \mathbf{x}(n) = \mathbf{0}$.

We first whiten the mean removed data. The idea of whitening is to make the data's components be uncorrelated and make their variances equal unity. The whitening processing is carried out as follows: $\tilde{\mathbf{x}}(n) = \mathbf{V}\mathbf{x}(n)$. $\mathbf{V} = \Gamma^{-1/2}$ is the whitening matrix, where $\Gamma = E\{\mathbf{x}\mathbf{x}^T\} = (1/N)\mathbf{X}\mathbf{X}^T$ is the covariance matrix. The spectral signatures of the interesting targets $\mathbf{d}_i$ should be dealt by $\tilde{\mathbf{d}}_i = \mathbf{V}\mathbf{d}_i$ ($i = 1, \ldots, m$). After the whitening processing, the covariance of the whitened data $\tilde{\mathbf{x}}$ is:

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = E\{\mathbf{V}\mathbf{x}\mathbf{x}^T\mathbf{V}^T\} = \mathbf{V}E\{\mathbf{x}\mathbf{x}^T\}\mathbf{V}^T = \Gamma^{-1/2}\Gamma(\Gamma^{-1/2})^T = \mathbf{I} \quad (1)$$

This denotes that the whitened data's components are uncorrelated and their variances equal unity.

After the whitening processing, we need to find an optimal projection vector, so targets can be detected in the projection space. In real hyperspectral images, targets of interest usually have a small population and occupy only a few pixels. In this case, targets of interest usually do not follow Gaussian distribution, and non-Gaussianity statistics could capture their spectral statistical characteristics better. We assume spectra of target pixels follow non-Gaussian distribution, and use non-Gaussianity statistics to characterize the spectral statistical characteristics of targets. The non-Gaussianity which can be measured by the non-Gaussianity statistics is mainly caused by target pixels. Thus, we can find an optimal projection vector $\mathbf{w} = [w_1, \ldots, w_L]^T$ by minimizing the non-Gaussianity statistics of the projection data $\mathbf{w}^T\tilde{\mathbf{x}}$. Because the non-Gaussianity is mainly caused by target pixels, in the projection space, the projections of spectra of background $\mathbf{w}^T\tilde{\mathbf{x}}_b$ and the projections of spectra of targets $\mathbf{w}^T\tilde{\mathbf{x}}_t$ will have different

values, where $\tilde{\mathbf{x}}_b$ and $\tilde{\mathbf{x}}_t$ denote the whitened spectra of background and targets, respectively. So targets and background can be separated in the projection space effectively. In order to make the algorithm stable, we add a constraint $\|\mathbf{w}\| = 1$. We hope that the projections of spectra of background $\mathbf{w}^T\tilde{\mathbf{x}}_b$ have small values, while the projections of spectra of targets $\mathbf{w}^T\tilde{\mathbf{x}}_t$ have large values, which requires $\mathbf{w}$ should be close to $\tilde{\mathbf{x}}_t$. But we cannot obtain the values of $\tilde{\mathbf{x}}_t$. Thus, we use the prior knowledge of targets $\tilde{\mathbf{d}}_i$ and regularization terms $\|\mathbf{w}-\tilde{\mathbf{d}}_i\|^2$ to guarantee the similarity between the projection vector $\mathbf{w}$ and the $\tilde{\mathbf{d}}_i$.

The problem becomes a constrained minimization problem as follows:

$$\min \ \Psi(\mathbf{w}) = E\{G(\mathbf{w}^T\tilde{\mathbf{x}})\} + \sum_{i=1}^{m} \lambda_i \|\mathbf{w}-\tilde{\mathbf{d}}_i\|^2$$

$$\text{s.t.} \quad \|\mathbf{w}\| = 1 \quad (2)$$

where E denotes expectation, and $\lambda_1, \ldots, \lambda_m$ are m constants. $E\{G(\cdot)\}$ is the non-Gaussianity statistic. For example, a classical measure of non-Gaussianity is the kurtosis [18] defined by

$$kurt(y) = E\{y^4\} - 3\{E[y^2]\}^2 \quad (3)$$

If the kurtosis is chosen, we get

$$kurt(\mathbf{w}^T\tilde{\mathbf{x}}) = E\{(\mathbf{w}^T\tilde{\mathbf{x}})^4\} - 3\{E[(\mathbf{w}^T\tilde{\mathbf{x}})^2]\}^2$$
$$= E\{(\mathbf{w}^T\tilde{\mathbf{x}})^4\} - 3\{\mathbf{w}^T E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T]\mathbf{w}\}^2 \quad (4)$$

After the whitening processing, the $\tilde{\mathbf{x}}$ has zero-mean and the covariance matrix equals $\mathbf{I}$. Noting that $\|\mathbf{w}\| = 1$, we get $kurt(\mathbf{w}^T\tilde{\mathbf{x}}) = E\{(\mathbf{w}^T\tilde{\mathbf{x}})^4\} - 3$. Thus, minimizing the kurtosis of the projection data $y = \mathbf{w}^T\tilde{\mathbf{x}}$ is equivalent to minimizing the fourth moment $E\{y^4\}$, and at this time $G(y) = y^4$. Different functions can be chosen as $G(\cdot)$ to measure the non-Gaussianity. Besides $G(y) = y^4$, some other functions could be chosen, such as: $G(y) = y^3$ and $G(y) = \log \cosh y$.

There is no closed form solution of (2), and we can use a simple gradient descent method [19] to solve this optimization problem. The gradient of $\Psi(\mathbf{w})$ to $\mathbf{w}$ is obtained as:

$$\frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}} = E\{\tilde{\mathbf{x}}g(\mathbf{w}^T\tilde{\mathbf{x}})\} + \sum_{i=1}^{m} 2\lambda_i(\mathbf{w}-\tilde{\mathbf{d}}_i) \quad (5)$$

where the function g is the derivative of G.

Thus, we have the following gradient descent iterations:

$$\mathbf{w} \leftarrow \mathbf{w} - \mu \frac{\partial \Psi(\mathbf{w})}{\partial \mathbf{w}} \quad (6)$$

$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\| \quad (7)$$

where μ is a learning rate. After the optimal projection vector $\mathbf{w}$ is obtained, we project each pixel's whitened spectrum $\tilde{\mathbf{x}}$ and get the detection result $\mathbf{w}^T\tilde{\mathbf{x}}$. The regularized non-Gaussianity based multiple-target detector (RNGMD) is obtained as follows:

Algorithm Outline: RNGMD

(1) Center the data to make the mean zero and whiten the data to give $\tilde{\mathbf{x}}(n)$. Initialize $\mathbf{w}$, such as $\mathbf{w} = [1, 0, 0, \ldots]^T$. Choose suitable constants $\lambda_1, \ldots, \lambda_m$ and a suitable learning rate μ. In this paper, $\lambda_1, \ldots, \lambda_m$ are set to 1, and μ is set to 0.001.

(2) Update the weight vector by

$$\mathbf{w} \leftarrow \mathbf{w} - \mu\{E[\tilde{\mathbf{x}}g(\mathbf{w}^T\tilde{\mathbf{x}})] + \sum_{i=1}^{m} 2\lambda_i(\mathbf{w}-\tilde{\mathbf{d}}_i)\} \quad (8)$$

$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\| \quad (9)$$

(3) Set a convergence criteria: if $\|\mathbf{w}-\mathbf{w}_{old}\| < 10^{-4}$, stop. If not converged, go back to step (2).

(4) Project the $\tilde{\mathbf{x}}(n)$ to $\mathbf{w}$, and get the detection result $\mathbf{w}^T\tilde{\mathbf{x}}(n)$.

The larger the detection result $\mathbf{w}^T\tilde{\mathbf{x}}(n)$ of a pixel is, the more likely targets are present in the pixel. Furthermore, a threshold η can be set, if $\mathbf{w}^T\tilde{\mathbf{x}}(n) > \eta$, we determine that targets are present in this pixel, else if $\mathbf{w}^T\tilde{\mathbf{x}}(n) < \eta$, we determine that targets are absent in this pixel.

Then we give the relationship between the RNGMD and the famous multiple-target detection algorithm: MTCEM. The MTCEM also finds an optimal projection vector $\mathbf{w}$ through solving the following optimization problem:

$$\min E\{(\mathbf{w}^T\mathbf{x})^2\}$$
$$\text{s.t.} \quad \mathbf{w}^T\mathbf{d}_i = 1 \quad i = 1, \ldots, m \tag{10}$$

By using the Lagrange multiplier method [19], (10) can be turned into an unconstrained minimization problem as follows:

$$\min E\{(\mathbf{w}^T\mathbf{x})^2\} - \sum_{i=1}^m \gamma_i \mathbf{w}^T\mathbf{d}_i \tag{11}$$

where γ_i $(i = 1, \ldots, m)$ are the Lagrange multipliers. On the other hand, noticing the constraint $\|\mathbf{w}\| = 1$, the objective function $\Psi(\mathbf{w})$ of the RNGMD in (2) can be rewritten as:

$$\Psi(\mathbf{w}) = E\{G(\mathbf{w}^T\tilde{\mathbf{x}})\} + \sum_{i=1}^m \lambda_i \|\mathbf{w} - \tilde{\mathbf{d}}_i\|^2$$
$$= E\{G(\mathbf{w}^T\tilde{\mathbf{x}})\} + \sum_{i=1}^m \lambda_i \mathbf{w}^T\mathbf{w} - \sum_{i=1}^m 2\lambda_i \mathbf{w}^T\tilde{\mathbf{d}}_i + \sum_{i=1}^m \lambda_i \tilde{\mathbf{d}}^T\tilde{\mathbf{d}}_i$$
$$= E\{G(\mathbf{w}^T\tilde{\mathbf{x}})\} + \sum_{i=1}^m \lambda_i - \sum_{i=1}^m \lambda_i' \mathbf{w}^T\tilde{\mathbf{d}}_i + \sum_{i=1}^m \lambda_i \tilde{\mathbf{d}}_i^T\tilde{\mathbf{d}}_i \tag{12}$$

where $\lambda_i' = 2\lambda_i$ $(i = 1, \ldots, m)$. For given hyperspectral images, λ_i', λ_i, and $\tilde{\mathbf{d}}_i^T\tilde{\mathbf{d}}_i$ $(i = 1, \ldots, m)$ are all constants, and have no influence on the minimization problem. Thus, the minimization problem of the RNGMD turns into

$$\min \Psi(\mathbf{w}) = E\{G(\mathbf{w}^T\tilde{\mathbf{x}})\} - \sum_{i=1}^m \lambda_i' \mathbf{w}^T\tilde{\mathbf{d}}_i$$
$$\text{s.t.} \quad \|\mathbf{w}\| = 1 \tag{13}$$

Comparing (13) with (11), we find if we choose $G(y) = y^2$, and do not use the whitened data, and discard the constraint $\|\mathbf{w}\| = 1$, (13) turns into (11). If we choose $G(y) = y^2$, we actually use the second-order statistic. Thus, the MTCEM can be seen as a special case of the RNGMD when the second-order statistic is used, and the whitening processing is omitted. In fact, the non-Gaussianity statistics could describe the non-Gaussian data more effectively than the second-order statistics. The whitening processing can eliminate the influence of second-order statistics, and the constraint $\|\mathbf{w}\| = 1$ can make the algorithm more stabler in the iterative process. The MTCEM becomes the CEM when one target source is detected. Thus, the CEM can also be seen as a special case of the RNGMD.

3. Stability of the RNGMD

In this section, we give the stability of the RNGMD. We have the following theorem:

Theorem 1. Assume that spectra of the hyperspectral image follow the linear spectral mixing model: $\mathbf{x} = \mathbf{Su} = \sum_{i=1}^M \mathbf{s}_i u_i$, where $\mathbf{x}$ is the mean removed spectrum. $\mathbf{s}_1, \ldots, \mathbf{s}_M$ are M deterministic spectra named endmembers, and $u_1, \ldots, u_M$ are the corresponding abundances. Furthermore, we assume M is equal to the spectral band number L. Also, we assume abundances are independent with each other, and assume $E\{\mathbf{u}\} = \mathbf{0}$, $E\{\mathbf{uu}^T\} = \mathbf{I}$. The whitened spectrum is

$\tilde{\mathbf{x}} = \mathbf{VSu}$, where $\mathbf{V}$ is the whitened matrix. Denote $\mathbf{A} = \mathbf{VS}$, and $\mathbf{a}_1, \ldots, \mathbf{a}_M$ are the columns of matrix $\mathbf{A}$. $\tilde{\mathbf{d}}_i$ $(i = 1, \ldots, m)$ are assumed to locate in the orthogonal subspace of $span\{\mathbf{a}_l(l = 1, \ldots, M, l \neq k)\}$, where $\tilde{\mathbf{d}}_i$ $(i = 1, \ldots, m)$ are the whitened spectral signatures of the interesting targets. Then with the constraint $\|\mathbf{w}\| = 1$, the local minima of $\Psi(\mathbf{w})$ in (2) include the k th column of the matrix $\mathbf{A}$ such that

$$E\{g'(u_k) - u_k g(u_k)\} + 2\sum_{i=1}^m \lambda_i \mathbf{a}_k^T\tilde{\mathbf{d}}_i > 0 \tag{14}$$

where the function g is the derivative of G, and g' is the derivative of g.

Proof. We only prove the case that $k = 1$, the proofs of the other cases ($k \neq 1$) are similar. We assume spectra of the hyperspectral image follow the linear spectral mixing model [2,20]: $\mathbf{x} = \mathbf{Su} = \sum_{i=1}^M \mathbf{s}_i u_i$, where $\mathbf{x}$ is the mean removed spectrum. For the linear spectral mixing model, the spectrum $\mathbf{x}$ is assumed to be the linear mixture of M deterministic spectra $\mathbf{s}_1, \ldots, \mathbf{s}_M$ named endmembers. $u_1, \ldots, u_M$ are the corresponding abundances which are assumed to be independent with each other.

After the whitening processing, we have $\tilde{\mathbf{x}} = \mathbf{Vx}$, where $\mathbf{V}$ is the whitened matrix, and

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{VS}E\{\mathbf{uu}^T\}\mathbf{S}^T\mathbf{V}^T = (\mathbf{VS})E\{\mathbf{uu}^T\}(\mathbf{VS})^T = \mathbf{I} \tag{15}$$

Under the assumption $E\{\mathbf{uu}^T\} = \mathbf{I}$, we denote $\mathbf{A} = \mathbf{VS}$ and have $\mathbf{AA}^T = \mathbf{I}$. Noting that $M = L$, A is a square matrix. Thus, $\mathbf{A}$ is an orthogonal matrix.

We make the transform $\mathbf{z} = \mathbf{A}^T\mathbf{w}$. Noticing that $\mathbf{A}$ is an orthogonal matrix, (2) becomes

$$\min \Psi_2(\mathbf{z}) = E\{G(\mathbf{z}^T\mathbf{u})\} + \sum_{i=1}^m \lambda_i \|\mathbf{Az} - \tilde{\mathbf{d}}_i\|^2$$
$$\text{s.t.} \quad \|\mathbf{z}\| = 1 \tag{16}$$

The gradient and the Hessian of $\Psi(\mathbf{z})$ are

$$\nabla\Psi_2(\mathbf{z}) = E\{g(\mathbf{z}^T\mathbf{u})\mathbf{u}\} + 2\sum_{i=1}^m \lambda_i \mathbf{z} - 2\sum_{i=1}^m \lambda_i \mathbf{A}^T\tilde{\mathbf{d}}_i \tag{17}$$

$$\nabla^2\Psi_2(\mathbf{z}) = E\{\mathbf{uu}^T g'(\mathbf{z}^T\mathbf{u})\} + 2\sum_{i=1}^m \lambda_i \mathbf{I} \tag{18}$$

We analyze the stability of the point $\mathbf{z} = \mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0, 0, \ldots)^T$. Calculating the gradient and the Hessian at point $\mathbf{z} = \mathbf{e}_1$, and utilizing the independence of the u_i, we obtain

$$\nabla\Psi_2(\mathbf{e}_1) = \mathbf{e}_1 E\left\{u_1 g(u_1) + 2\sum_{i=1}^m \lambda_i\right\} - 2\sum_{i=1}^m \lambda_i \mathbf{A}^T\tilde{\mathbf{d}}_i \tag{19}$$

$$\nabla^2\Psi_2(\mathbf{e}_1) = diag\left\{E[u_1^2 g'(u_1)] + 2\sum_{i=1}^m \lambda_i, E[g'(u_1)] + 2\sum_{i=1}^m \lambda_i, \right.$$
$$\left. E[g'(u_1)] + 2\sum_{i=1}^m \lambda_i, \cdots\right\} \tag{20}$$

Making a small perturbation $\boldsymbol{\varepsilon}$, we obtain

$$\Psi_2(\mathbf{e}_1 + \boldsymbol{\varepsilon}) = \Psi_2(\mathbf{e}_1) + \boldsymbol{\varepsilon}^T\nabla\Psi_2(\mathbf{e}_1) + \frac{1}{2}\boldsymbol{\varepsilon}^T\nabla^2\Psi_2(\mathbf{e}_1)\boldsymbol{\varepsilon} + o(\|\boldsymbol{\varepsilon}\|)^2$$
$$= \Psi_2(\mathbf{e}_1) + \varepsilon_1 E\left\{u_1 g(u_1) + 2\sum_{i=1}^m \lambda_i\right\} - 2\sum_{k=1}^M \sum_{i=1}^m \varepsilon_k \lambda_i \mathbf{a}_k^T\tilde{\mathbf{d}}_i$$
$$+ \frac{1}{2}\left\{E\left[u_1^2 g'(u_1) + 2\sum_{i=1}^m \lambda_i\right]\varepsilon_1^2 + E\left[g'(u_1) + 2\sum_{i=1}^m \lambda_i\right]\sum_{j \neq 1}\varepsilon_j^2\right\}$$
$$+ o(\|\boldsymbol{\varepsilon}\|^2) \tag{21}$$

where $\mathbf{a}_1,\ldots,\mathbf{a}_M$ are the columns of matrix $\mathbf{A}$. Because of the constraint $\|\mathbf{z}\|=1$, we obtain $\varepsilon_1 = \sqrt{1-\sum_{j\neq 1}\varepsilon_j^2}-1$. We notice that $\sqrt{1-\beta}=1-\beta/2+o(\beta)$. Thus, the term of order ε_1^2 in (21) is $o(\|\varepsilon\|^2)$, i.e., of higher order, and can be neglected. Using the first-order approximation for ε_1, we obtain $\varepsilon_1 = -\sum_{j>1}\varepsilon_j^2/2+o(\|\varepsilon\|^2)$. We assume that $\tilde{\mathbf{d}}_i$ $(i=1,\ldots,m)$ locate in the orthogonal subspace of $span\{\mathbf{a}_2,\ldots,\mathbf{a}_M\}$, which means $\tilde{\mathbf{d}}_i$ $(i=1,\ldots,m)$ are orthogonal with $\mathbf{a}_2,\ldots,\mathbf{a}_M$. This assumption is reasonable, because this assumption means $\mathbf{a}_2,\ldots,\mathbf{a}_M$ denote these undesired whitened spectral signatures, and the whitened spectral signatures of the interesting targets locate in the orthogonal subspace of $span\{\mathbf{a}_2,\ldots,\mathbf{a}_M\}$. Finally, we obtain

$$\Psi_2(\mathbf{e}_1+\varepsilon) = \Psi_2(\mathbf{e}_1)+\frac{1}{2}\left\{E[g'(u_1)-u_1 g(u_1)]+2\sum_{i=1}^{m}\lambda_i \mathbf{a}_1^T \tilde{\mathbf{d}}_i\right\}\sum_{j\neq 1}\varepsilon_j^2$$
$$+o(\|\varepsilon\|^2) \tag{22}$$

which proves $\mathbf{z}=\mathbf{e}_1$ or $\mathbf{w}=\mathbf{a}_1$ is a minimum under the conditions of the theorem. □

4. Experimental results

In this section, we used two synthetic hyperspectral images and two real hyperspectral images to do multiple-target detection experiments. The proposed RNGMD was only compared with multiple-target detection methods including the TCIMF method and the three CEM based methods: the WTACEM, the MTCEM, and the SCEM. In fact, the WTACEM and the SCEM all apply several CEM detectors in parallel, one for each target, and use the composite results of these parallel CEM detectors to detect multiple targets, but the experimental results will show that these parallel detectors based methods do not work well.

Receiver operating characteristic (ROC) [2,21] curves were used to evaluate algorithms quantitatively. By taking a threshold, the detection result can be transformed to a binary image, where value 1 represents targets are present in the pixel, and value 0 represents targets are absent. Based on the ground truth, by taking all possible thresholds, the ROC curve can plot the varying relationship between the probability of detection (Pd) and the false alarm rate (Fa), which are defined as:

$$Pd = \frac{N_d}{N_t}, \quad Fa = \frac{N_f}{N_b} \tag{23}$$

where N_d is the number of correct detected target pixels in the binary image, N_t is the total number of truth target pixels, N_f is the number of false alarm pixels in the binary image, and N_b is the total number of truth background pixels. A higher ROC curve which means a higher detection probability in the same false alarm rate represents a better detection algorithm.

4.1. Synthetic image experiment 1

We used the method introduced by Chang et al. [22] to design synthetic hyperspectral images. Five pure pixels: Airplane (A), Grass (G), Roof (R), Parking apron (P), and Ground (Gd) were chosen from a real airborne visible/infrared imaging spectrometer (AVIRIS) hyperspectral image. The mean spectrum of the real AVIRIS image denoted by M was also calculated out. The real AVIRIS image has 224 original bands with the spectral range of 0.4 μm–2.5 μm. After removing the water absorption and low SNR bands, the experiments used spectra have 189 bands left. Fig. 1 shows spectra of A, G, R, P, Gd and M. We used these six kinds of spectra to design the synthetic image 1 with the size of 200×200 pixels. The targets were simulated by spectra of A, G, R, P and Gd. The background was simulated by M. Fig. 2 shows the first band of the synthetic image 1.
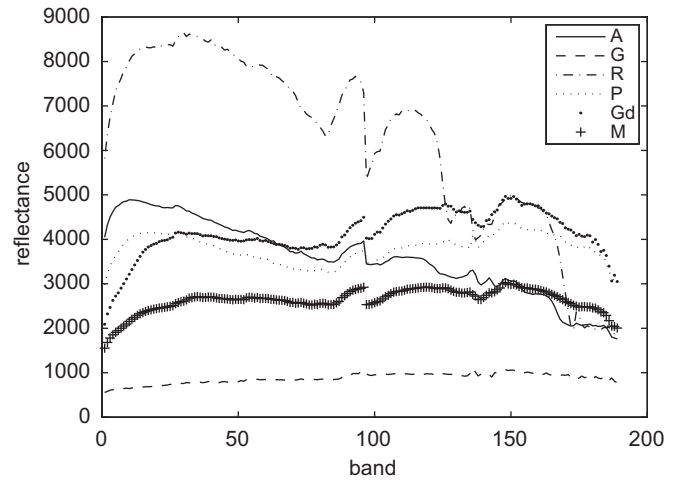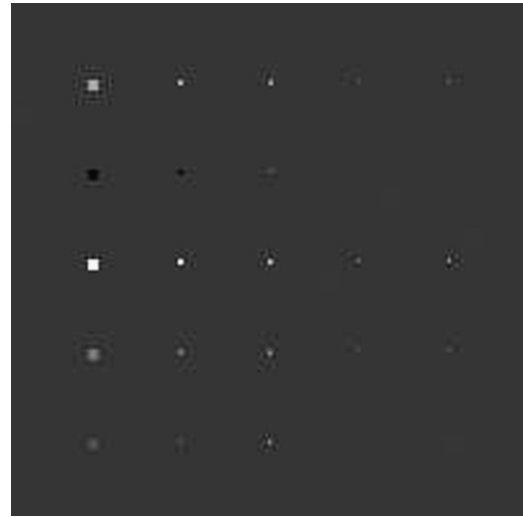

Fig. 1. Spectra of A, G, R, P, Gd and M.

Fig. 2. The first band of the synthetic image 1. In columns 1 and 2, from top to bottom, there are five 4×4 and 2×2 pure-pixel panels simulated by spectra of A, G, R, P and Gd, respectively. Five 2×2 mix-pixel panels locate in column 3, and five 1×1 sub-pixel panels locate in column 4 and column 5, respectively.

In the synthetic image 1, there are 25 target panels arranged in a 5×5 matrix, including 10 pure-pixel target panels, 5 mix-pixel target panels and 10 sub-pixel target panels. The total number of target pixels is 130. An additive white Gaussian noise was added to the synthetic image 1 to achieve a 50 dB signal-to-noise ratio (SNR).

The synthetic image 1 was used to detect the five kinds of targets: A, G, R, P, and Gd. Fig. 3 shows every band's kurtosis of the spectra of targets in the synthetic image 1. The kurtosis is a classical measure of non-Gaussianity. For Gaussian data, the kurtosis is zero, while for non-Gaussian data, the kurtosis is nonzero. From Fig. 3, we can clearly see that the spectra of targets indeed do not follow Gaussian distribution in the synthetic image 1. $G(y)=y^4$ known as the kurtosis was used as the non-Gaussianity measured function in the RNGMD. Spectra of A, G, R, P, and Gd as shown in Fig. 1 were used as the prior knowledge of spectral signatures of interesting targets in all the experimental algorithms. The M was used as the prior knowledge of undesired signature in the TCIMF. Fig. 4(a)–(e) shows the detection results of the RNGMD, the TCIMF, the WTACEM, the MTCEM, and the SCEM, respectively. ROC curves of different algorithms are shown in Fig. 5, where ROC curves of different algorithms are overlapped and all locate in 1. This means no matter what the false alarm rates are, the probabilities of
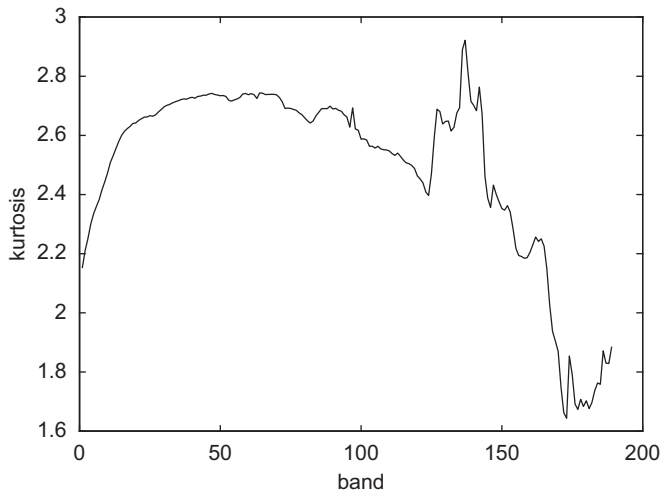
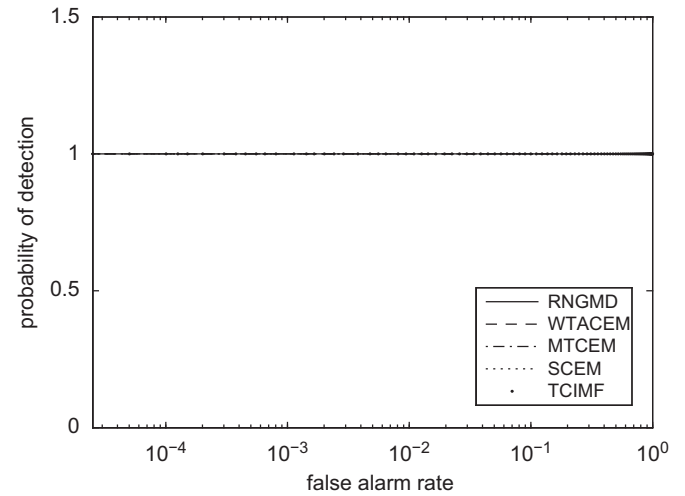**Fig. 3.** Every band's kurtosis of the spectra of targets in the synthetic image 1.

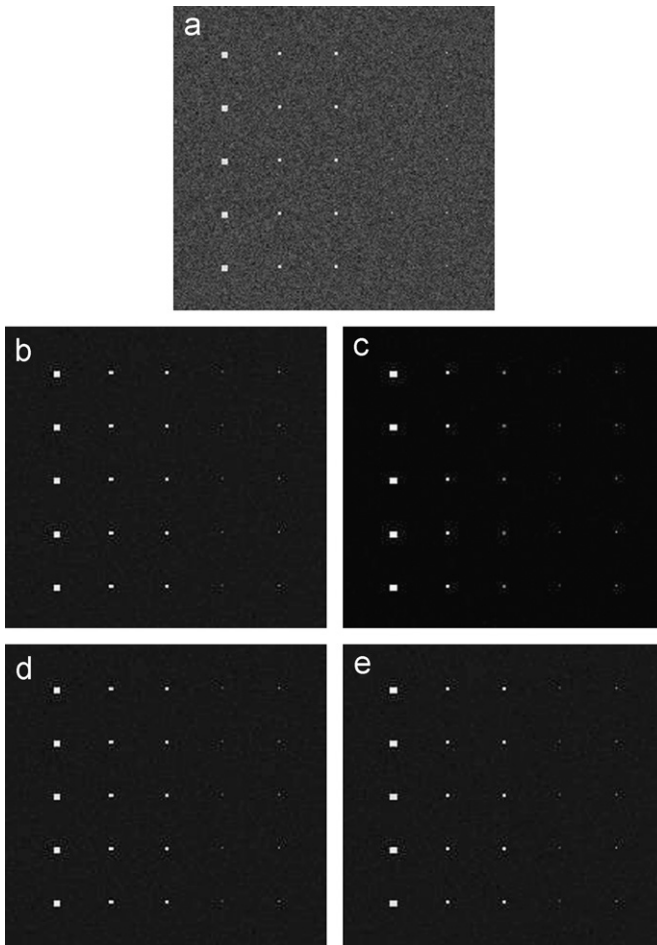**Fig. 5.** ROC curves of different algorithms for the synthetic image 1.

Fig. 4. (a) The detection result of the RNGMD for the synthetic image 1. (b) The detection result of the TCIMF for the synthetic image 1. (c) The detection result of the WTACEM for the synthetic image 1. (d) The detection result of the MTCEM for the synthetic image 1. (e) The detection result of the SCEM for the synthetic image 1.

Fig. 6. The first band of the synthetic image 2.

4.2. Synthetic image experiment 2

The synthetic image 2 is similar to the synthetic image 1. The only difference was that the synthetic image 2 was corrupted by a stronger additive white Gaussian noise to achieve 10 dB SNR. Fig. 6 shows the first band of the synthetic image 2. The synthetic image 2 was also used to detect the five kinds of targets: A, G, R, P, and Gd. Experimental algorithms' parameters were the same as those used in the synthetic image experiment 1. Fig. 7(a)–(e) shows the detection results of the RNGMD, the TCIMF, the WTACEM, the MTCEM, and the SCEM, respectively. From Fig. 7, we can see that the RNGMD can detect targets well, while the other experimental algorithms nearly miss all targets. The results denote that under the circumstance of low SNR, the RNGMD performs better, and the other four algorithms seem to be sensitive to noise. ROC curves of different algorithms are shown in Fig. 8, where the ROC curve of the RNGMD is much higher than those of the other experimental algorithms when the false alarm rate is lower than about 0.3. The ROC curve of the RNGMD is a little lower than those of the other experimental algorithms in high false alarm rates. But the RNGMD has better detection result on the whole.

detection are always 1. Fig. 5 indicates that the detection results of different experimental algorithms are equivalent in the quantitative sense for the synthetic image 1. From Figs. 4 and 5, we can see all algorithms can detect targets well. This is due to the fact that the SNR is very high (the SNR is 50 dB), which means that there is nearly no noise in the image.
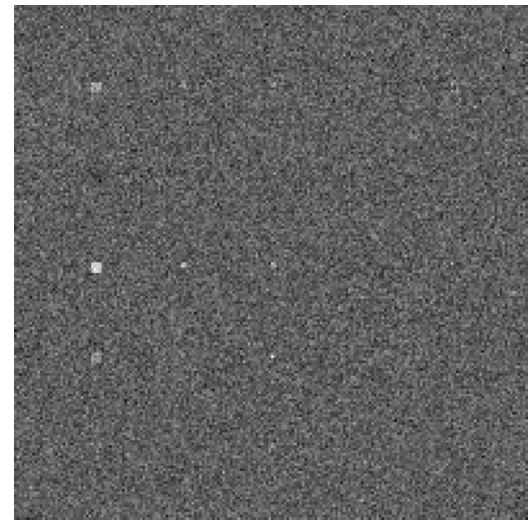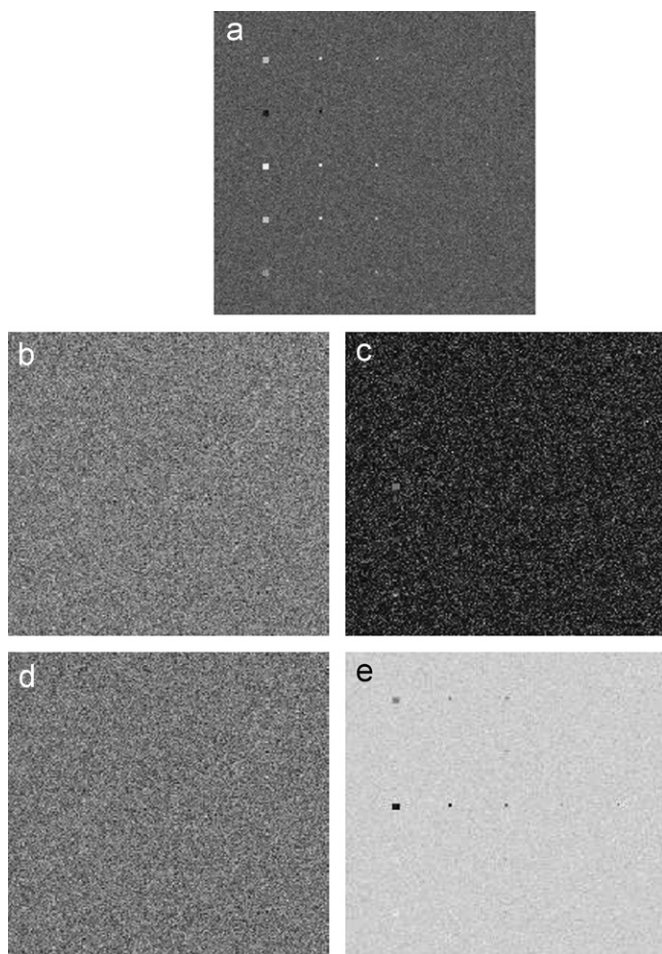
Fig. 7. (a) The detection result of the RNGMD for the synthetic image 2. (b) The detection result of the TCIMF for the synthetic image 2. (c) The detection result of the WTACEM for the synthetic image 2. (d) The detection result of the MTCEM for the synthetic image 2. (e) The detection result of the SCEM for the synthetic image 2.
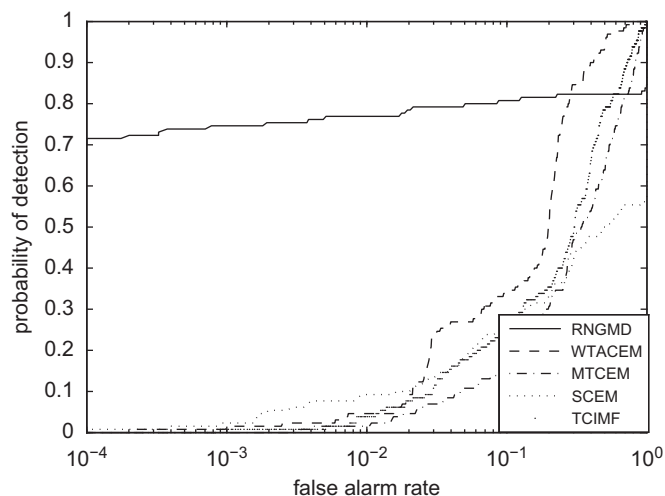

Fig. 8. ROC curves of different algorithms for the synthetic image 2.

4.3. Real image experiment 1

In this section, we used a real hyperspectral image collected by the AVIRIS sensor to do experiments. The AVIRIS sensor collects spectral data in 224 bands, and the spectral range is 0.4 μm–2.5 μm. We have removed the water absorption and low SNR bands, and 189 available bands were left. Fig. 9 shows the first band of the AVIRIS image. The image has 200×300 pixels, and the scene is a part of the San Diego airport, San Diego, USA. There are two kinds of airplanes in the image. The first kind of
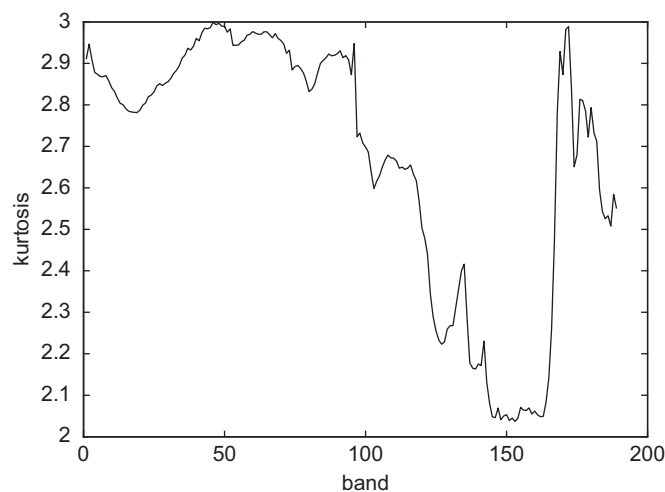

Fig. 9. The first band of the AVIRIS image.

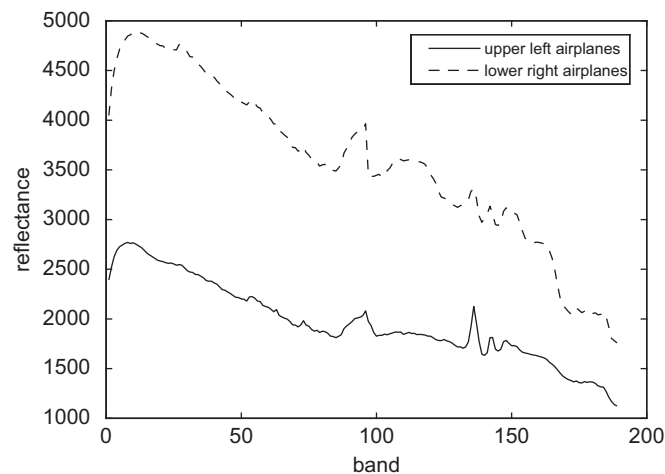**Fig. 10.** Every band's kurtosis of the spectra of targets in the AVIRIS image.

Fig. 11. The prior knowledge of spectral signatures of the two kinds of airplanes in the AVIRIS image.

three airplanes locate in the upper left corner of the image, and the second kind of three airplanes locate in the lower right corner of the image. We used this AVIRIS image to detect the two kinds of airplanes. The total number of target pixels is 204. Fig. 10
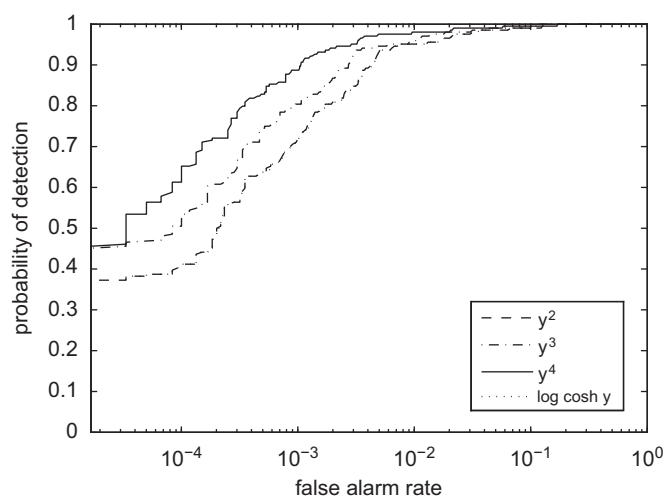

Fig. 12. ROC curves of different non-Gaussianity measured functions for the AVIRIS image.

shows every band's kurtosis of the spectra of targets in the AVIRIS image. Fig. 10 indicates that the spectra of targets do not follow Gaussian distribution once again. Spectra shown in Fig. 11 were used as the prior knowledge of spectral signatures of the two kinds of airplanes in all the experimental algorithms.

In order to compare the detection results of different non-Gaussianity measured functions, we chose three different non-Gaussianity measured functions: $G(y) = y^3$, $G(y) = y^4$, and $G(y) = \log \cosh y$. Although $G(y) = y^2$ is not a non-Gaussianity measured function, we also added it for comparison. Fig. 12 shows the ROC curves of different non-Gaussianity measured functions, where the ROC curves of the $G(y) = y^2$ and the $G(y) = \log \cosh y$ are nearly overlapped. Fig. 12 shows that $G(y) = y^4$ performs best, which demonstrates that the kurtosis is a good choice of the non-Gaussianity statistics. Thus, in the following experiments, $G(y) = y^4$ was chosen as the non-Gaussianity measured function. Because the TCIMF eliminates undesired signatures, the TCIMF requires the prior spectral knowledge of undesired signatures. Unfortunately, for real hyperspectral images, we cannot obtain this knowledge directly, so we cannot implement the TCIMF for real hyperspectral images. Fig. 13(a)–(d) shows the detection results of the RNGMD, the WTACEM, the MTCEM, and the SCEM, respectively. We transformed the detection results of different algorithms to binary images as shown in Fig. 14. The thresholds were properly selected so that the false alarm rates in Fig. 14 were 0.001. From Fig. 14, we can see the RNGMD could detect the targets more effectively. Fig. 15 shows
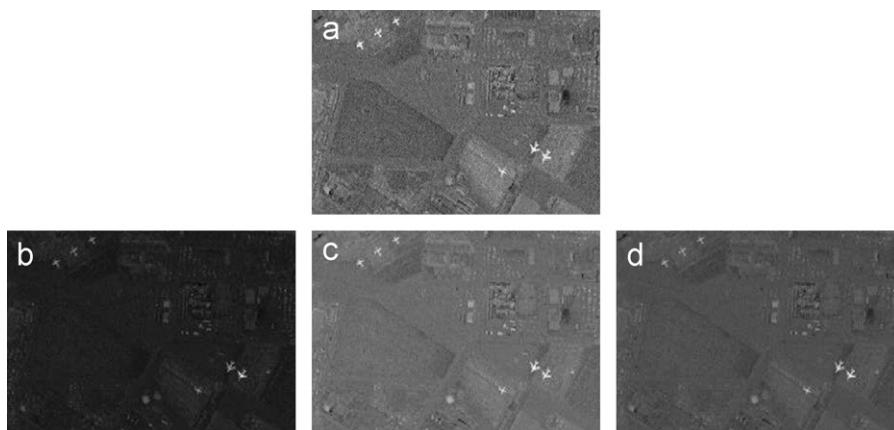

Fig. 13. (a) The detection result of the RNGMD for the AVIRIS image. (b) The detection result of the WTACEM for the AVIRIS image. (c) The detection result of the MTCEM for the AVIRIS image. (d) The detection result of the SCEM for the AVIRIS image.
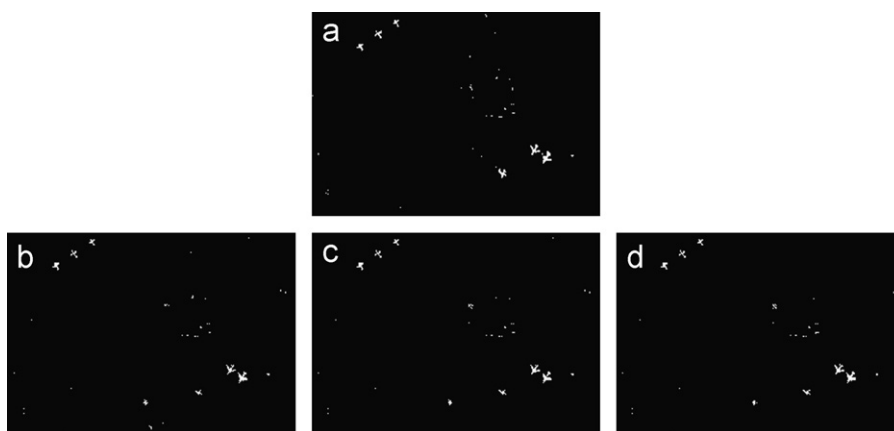

Fig. 14. (a) The detection result of the RNGMD in the form of binary image for the AVIRIS image. (b) The detection result of the WTACEM in the form of binary image for the AVIRIS image. (c) The detection result of the MTCEM in the form of binary image for the AVIRIS image. (d) The detection result of the SCEM in the form of binary image for the AVIRIS image.

ROC curves of different algorithms. From Fig. 15, it can be found that in the same false alarm rate, the probability of detection of the RNGMD is always higher than those of the other experimental algorithms. Thus, the proposed RNGMD performs better than the other experimental algorithms. In the AVIRIS image, targets of interest only occupy a few pixels, and spectra of targets do not follow Gaussian distribution. The RNGMD uses non-Gaussianity statistics to describe statistical characteristics of targets, while the other experimental algorithms use the second-order statistics, which is suitable to describe Gaussian data. Thus, the RNGMD has better performance than other experimental algorithms.
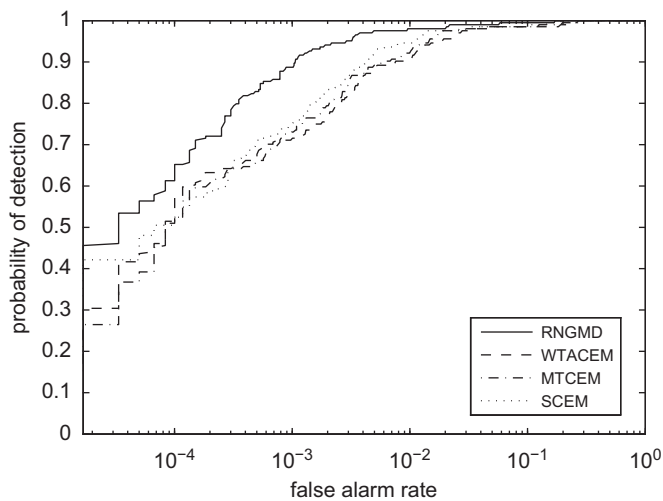
4.4. Real image experiment 2

In this section, a real hyperspectral image collected by the hyperspectral mapper (HyMap) sensor was used to do experiments. The HyMap sensor collects spectral data in 126 bands, and the spectral range is $0.45\mu m-2.5\mu m$. The HyMap image can be obtained from the internet [23,24]. The website provides two data sets, a self test set and a blind test set. Since the self test set has the ground truth of targets, it is suitable for evaluating algorithms. Accordingly, we used the self test set to do experiments. The HyMap images are available both in calibrated spectral radiance as well as in spectral reflectance after atmospheric compensation [23]. We used the reflectance data. The HyMap image has 280×800 pixels and 126 bands. The scene was the small town of Cooke City, Montana, USA. Fig. 16 shows the first band of the HyMap image. We used the HyMap data to detect a fabric panel target and a vehicle target. Both the fabric panel target and the vehicle target occupy only one pixel. Thus, we cannot identify them in the image with our naked eyes. Fig. 17 shows the true distribution named ground truth of the fabric panel target and the vehicle target. The ground truth can be obtained from the internet [24]. In Fig. 17, the right bright point is the position of the fabric panel target, and the left bright point is the position of the vehicle target. Spectra shown in Fig. 18 were used as the prior knowledge of spectral signatures of the fabric panel target and the vehicle target in all the experimental algorithms. The detection of these two targets is really difficult, because there are only two target pixels, while the total number of pixels in the HyMap image is 224 000.

Since $G(y)=y^4$ did best in the previous experiments, it was chosen as the non-Gaussianity measured function. Like the AVIRIS image experiments, the proposed RNGMD algorithm was compared with the WTACEM, the MTCEM, and the SCEM. Fig. 19(a)–(d) shows the detection results of the RNGMD, the WTACEM, the MTCEM, and the SCEM, respectively. As a quantitative analysis, Fig. 20 shows ROC

Fig. 15. ROC curves of different algorithms for the AVIRIS image.

Fig. 16. The first band of the HyMap image.

Fig. 17. The true distribution of the fabric panel target and the vehicle target in the HyMap image.
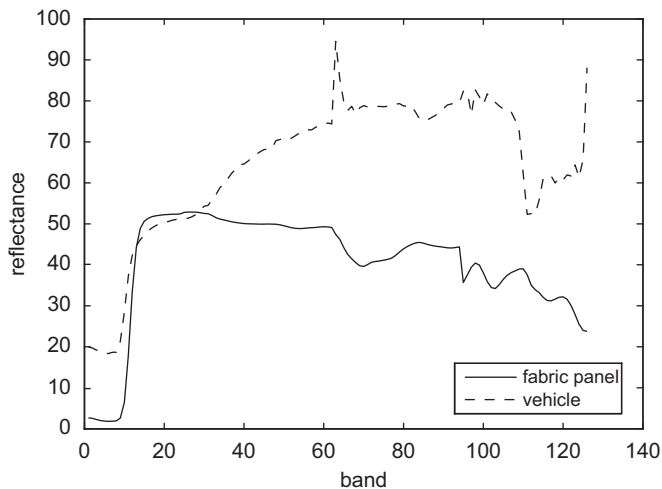
Fig. 18. The prior knowledge of spectral signatures of the fabric panel target and the vehicle target in the HyMap image.
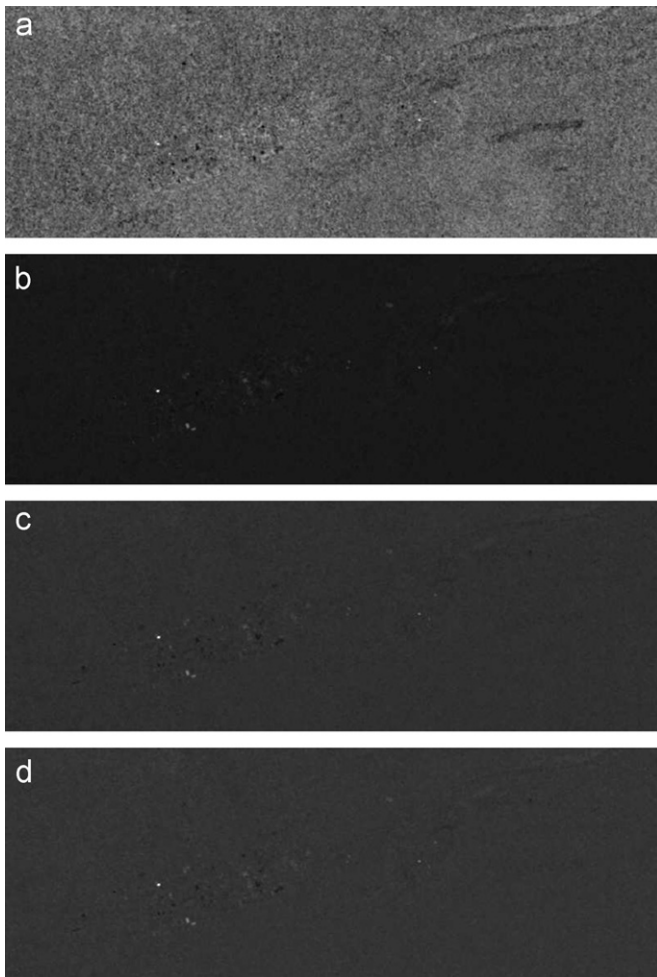

Fig. 19. (a) The detection result of the RNGMD for the HyMap image. (b) The detection result of the WTACEM for the HyMap image. (c) The detection result of the MTCEM for the HyMap image. (d) The detection result of the SCEM for the HyMap image.

Fig. 20. ROC curves of different algorithms for the HyMap image.

Table 1
The number of false alarms of different algorithms for the HyMap image.

Algorithms	RNGMD	WTACEM	MTCEM	SCEM
The number of false alarms	**712**	4379	80,832	57,535

alarm rate is 0, the detection probability of the RNGMD is 0.5. This is due to the fact that the fabric panel target located pixel has the largest RNGMD's output value. Thus, if we set a threshold, we can always detect the fabric panel target successfully. But for the other experimental algorithms, when the false alarm rates are 0, the detection probabilities are all 0. Because there are only two target pixels, and targets have no shape information, it is hard to compare detection results using binary images as Fig. 14. We could compare detection results in the following way. For different algorithms, we got the output values of the fabric panel target a_f and the output values of the vehicle target a_v, and let $a = \min\{a_f, a_v\}$. The number of background pixels whose output values were equal to or larger than a was counted. This number denotes the number of false alarms when the two kinds of targets are detected successfully. The smaller the number is the better performance the algorithm does. Table 1 gives the numbers of false alarms of different algorithms, where the bold fond is the smallest value. From Table 1, we can clearly see our proposed RNGMD has the smallest number of false alarms, and thus the best performance. The experimental results indicate that non-Gaussianity statistics could characterize targets more effectively than second-order statistics, and the RNGMD could behave better for this real hyperspectral image.

4.5. Experimental results analysis

Two synthetic images and two real images were used to do experiments. For the synthetic image 1, all algorithms can detect targets well, and ROC curves are very high. This is due to the fact that there is nearly no noise in the synthetic image 1. The synthetic image 2 was seriously corrupted by noise. The RNGMD could still detect the targets well, while the other experimental algorithms did not work well. The results demonstrate that the RNGMD is more robust to noise. For the real AVIRIS and HyMap hyperspectral images, the results also show that the RNGMD does best among the experimental algorithms. The experimental results indicate non-Gaussianity statistics indeed can characterize targets more effectively. We compared the performances of different non-Gaussianity statistics, and found the performance

curves of different algorithms. Because there are only two target pixels, the probability of detection can only be 0, 0.5, or 1 as shown in Fig. 20. The RNGMD has the highest ROC curve, and this indicate that the RNGMD has the best performance. Especially, even the false
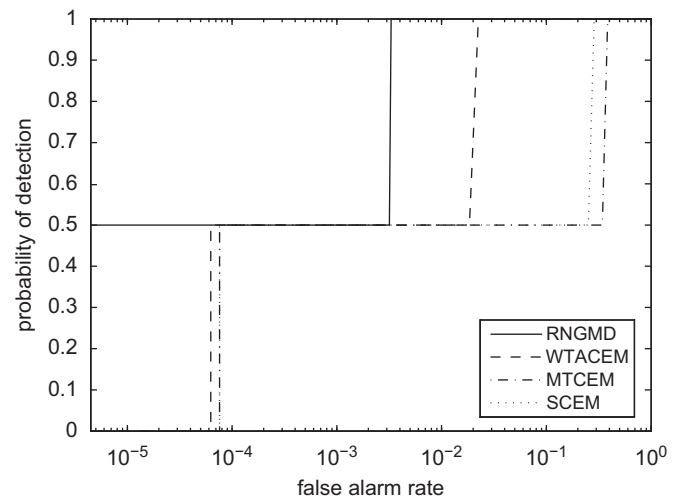
of the kurtosis was best. This proves that the kurtosis is an effective measure of non-Gaussianity. In fact, the kurtosis has been proved to be an effective measure of non-Gaussianity in other fields, such as in the independent component analysis field [18].

5. Conclusion

For hyperspectral images, most existing target detection algorithms could detect only one kind of target at a time. In this paper, we proposed a regularized non-Gaussianity based multiple-target detector for hyperspectral images. We used the non-Gaussianity statistics and regularization terms to find an optimal projection vector, so that the interesting targets can be detected in the projection space. In order to solve the optimal projection vector found optimization problem, we simply use the gradient descent method. We also analyze the stability of the proposed algorithm. Two synthetic hyperspectral images and two real hyperspectral images were used to evaluate and compare different algorithms. The experimental results show that the proposed algorithm is more effective than the other experimental algorithms.

References

[1] Manolakis D, Lockwood R, Cooley T, Jacobson J. Is there a best hyperspectral detection algorithm? In: Proceedings of SPIE, Orlando, USA; 2009.
[2] Manolakis D, Marden D, Shaw GA. Hyperspectral image processing for automatic target detection applications. Lincoln Laboratory Journal 2003;14(1): 79–116.
[3] Foy BR, Theiler J, Fraser AM. Decision boundaries in two dimensions for target detection in hyperspectral imagery. Optics Express 2009;17(20):17391–411.
[4] Gu Y, Liu Y, Zhang Y. A selective KPCA algorithm based on high-order statistics for anomaly detection in hyperspectral imagery. IEEE Geoscience and Remote Sensing Letters 2008;5(1):43–7.
[5] Manolakis D, Shaw G. Detection algorithms for hyperspectral imaging applications. IEEE Signal Processing Magazine 2002;19(1):29–43.
[6] Shaw GA, Burke HK. Spectral imaging for remote sensing. Lincoln Laboratory Journal 2003;14(1):3–28.
[7] Farrand WH, Harsanyi JC. Mapping the distribution of mine tailings in the coeur d'Alene River valley, Idaho, through the use of a constrained energy minimization technique. Remote Sensing of Environment 1997;59(1):64–76.
[8] Harsanyi JC. Detection and classification of subpixel spectral signatures in hyperspectral image sequences. PhD dissertation. Baltimore County: University of Maryland; 1993.
[9] Kay SM. Fundamentals of statistical signal processing. Englewood Cliffs, NJ, USA: Prentice Hall; 1998.
[10] Manolakis DG, Ingle Vk, Kogon SM. Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering and array processing. Boston, USA: McGraw-Hill; 2000.
[11] Jiao X, Chang CI. Kernel-based constrained energy minimization (K-CEM). In: Proceedings of SPIE, Orlando, USA; 2008.
[12] Kwon H, Nasrabadi NM. Kernel matched signal detectors for hyperspectral target detection. In: Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition—workshops, Washington, USA; 2005.
[13] Sakla W, Chan A, Ji J, Sakla A. An SVDD-based algorithm for target detection in hyperspectral imagery. IEEE Geoscience and Remote Sensing Letters 2011;8(2):384–8.
[14] Chang CI. Hyperspectral imaging: techniques for spectral detection and classification. New York, USA: Kluwer Academic/Plenum Publishers; 2003.
[15] Ren H, Du Q, Chang CI, Jensen JO. Comparison between constrained energy minimization based approaches for hyperspectral imagery. In: IEEE workshop on advances in techniques for analysis of remotely sensed data; 2003. p. 244–8.
[16] Chang CI, Jiao X, Wu CC, Du Y, Chang ML. A review of unsupervised spectral target analysis for hyperspectral imagery. EURASIP Journal on Advances in Signal Processing 2010; 2010; http://dx.doi.org/10.1155/2010/503752.
[17] Ren H, Du Q, Wang J, Chang CI, Jensen JO, Jensen JL. Automatic target recognition for hyperspectral imagery using high-order statistics. IEEE Transactions on Aerospace and Electronic Systems 2006;42(4):1372–85.
[18] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Networks 2000;13(4–5):411–30.
[19] Nocedal J, Wright SJ. Numerical optimization. New York, USA: Springer; 2000.
[20] Adams JB, Smith MO, Gillespie AR. Imaging spectroscopy: interpretation based on spectral mixture analysis. Cambridge, UK: Cambridge University Press; 1993.
[21] Chang CI. Multiparameter receiver operating characteristic analysis for signal detection and classification. IEEE Sensors Journal 2010;10(3):423–42.
[22] Chang YCC, Ren H, Chang CI, Rand RS. How to design synthetic images to validate and evaluate hyperspectral imaging algorithms. In: Proceedings of SPIE, Orlando, USA; 2008.
[23] Snyder D, Kerekes J, Fairweather I, Crabtree R, Shive J, Hager S. Development of a web-based application to evaluate target finding algorithms. In: Proceedings of IEEE international geoscience and remote sensing symposium, New York; 2008. p. II-915–8.
[24] [Online]. Available: ⟨http://dirs.cis.rit.edu/blindtest/⟩.