OUTLIER-ROBUST DIMENSION REDUCTION AND ITS IMPACT ON HYPERSPECTRAL ENDMEMBER EXTRACTION

Hao-En Huang[†], *Tsung-Han Chan*[†], *ArulMurugan Ambikapathi*[†], *Wing-Kin Ma*[*], *Chong-Yung Chi*[†]

[†]Institute of Communications Engineering, National Tsing Hua Univ., Hsinchu, Taiwan
[*]Department of Electronic Engineering, Chinese Univ. Hong Kong, Shatin, N.T., Hong Kong
E-mail: b23004705@hotmail.com; {thchan,aareul,wkma}@ieee.org; cychi@ee.nthu.edu.tw

ABSTRACT

Hyperspectral endmember extraction is a process to extract endmember signatures from the observed hyperspectral data of an area. The presence of outliers in the data has been proved to pose a serious problem in endmember extraction. In this paper, unlike conventional outlier detectors which may be sensitive to window settings, we propose a robust affine set fitting (RASF) algorithm for joint dimension reduction and outlier detection without any window setting. Given the number of endmembers in advance, the RASF algorithm is to find a data-representative affine set from the corrupted data, while making the effects of outliers minimum, in the least-squares error sense. The proposed RASF algorithm is then combined with Neyman-Pearson hypothesis testing, termed RASF-NP, to further estimate the number of outliers present in the data. Computer simulations demonstrate the efficacy of the proposed method, and its impact on existing endmember extraction algorithms.

Index Terms— Hyperspectral images, Robust dimension reduction, Endmember extraction

1. INTRODUCTION

In the past several years, endmember extraction using hyperspectral images has been widely investigated and proven to be valuable in many applications, including geology, hydrology, urban planning, geography, cadastral mapping, cartography, and military [1]. However, the presence of outliers in the hyperspectral data is inevitable in practice, and may seriously affect the analysis of hyperspectral data. The outliers are thought of as the pixels that appear to deviate markedly from the rest of the data. Two definitions of outlier pixels have been presented in [2, 3]. The first one refers to the pixels that provide constant or error readout, also called "dead" or "bad" pixels. Possible causes include detector failure, errors during data transfer, and improper data correction. The second one refers to the pixels that have rather different spectral signatures from the background representative. These pixels are also commonly called targets in the domain of hyperspectral anomaly detection.

Present outlier detection (OD) methods conceptually detect the outliers based on some sort of distance measure between outliers and background representative. The RX algorithm [4], known as a benchmark OD algorithm, assumes that all the background pixel vectors have the same multivariate normal distribution, and uses a sliding window scheme to compute the background covariance matrix. In the sliding window, the centered pixel and the rest of pixels correspond to the tested target and background, respectively. Since some probable outliers may be taken in the background region of the window, the calculation of the background covariance matrix is not accurate anymore, leading to performance degradation of RX algorithm. To properly mitigate this problem, random-selection-based anomaly detector (RSAD) [5] is reported to robustly compute background information, intending to take as few outliers involved in background as possible. But it may spend much more computation time.

In this paper, we focus not only on the OD problem, but also on the dimension reduction. We develop a robust affine set fitting (RASF) algorithm, a robust version of the affine set fitting (ASF) [6], for joint dimension reduction and outlier detection. Assuming the number of endmembers and outliers are known in advance, RASF is to find a contamination-free, data-representative affine set from the corrupted data, while minimizing the outlier effects in the least-squares error sense. The proposed RASF algorithm does not rely on any sliding window setting, and is implemented by alternating optimization. Furthermore, we incorporate the estimation of the number of outliers in the RASF algorithm, by Neyman-Pearson hypothesis testing; the resulting algorithm will be called RASF-NP. Simulations will show the effectiveness of the proposed RASF and RASF-NP algorithms, and comparison of the RASF-NP algorithm with RSAD method [5], and the impact of the RASF-NP algorithm on some existing benchmark endmember extraction algorithms (EEAs).

Notations: $\mathbb{R}^N$ and $\mathbb{R}^{M\times N}$ denote set of real $N\times 1$ vectors and set of real $M\times N$ matrices, respectively; $\mathbf{0}$ is the all-zero vector with proper dimension; $\mathbf{I}_N$ represents $N\times N$ identity matrix; "$\|\cdot\|$" stands for Euclidean norm; $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathbf{P}_{\mathbf{C}}^{\perp}$ is the orthogonal complement projector of matrix $\mathbf{C}$; $\lceil\cdot\rceil$ denotes the ceiling function.

2. PROBLEM STATEMENT AND ASSUMPTIONS

Consider an $M\times N$ linear spectral mixing model [7]:

$$\begin{aligned}\mathbf{y}[n] &= \mathbf{A}\mathbf{s}[n]+\mathbf{w}[n]+\mathbf{z}[n],\\ &= \mathbf{x}[n]+\mathbf{w}[n]+\mathbf{z}[n],\ n=1,\ldots,L,\end{aligned}\tag{1}$$

where $\mathbf{y}[n]=[\,y_1[n],\ldots,y_M[n]\,]^T\in\mathbb{R}^M$ is the nth observed pixel vector comprising M spectral bands, $\mathbf{x}[n]\triangleq\mathbf{A}\mathbf{s}[n]$ is the noise and outlier free counterpart, in which $\mathbf{A}=[\,\mathbf{a}_1,\ldots,\mathbf{a}_N\,]\in\mathbb{R}^{M\times N}$ denotes the signature matrix whose ith column vector $\mathbf{a}_i$ is the ith endmember signature (or simply, endmember), $\mathbf{s}[n]=[\,s_1[n],\ldots,s_N[n]\,]^T\in\mathbb{R}^N$ is the nth abundance vector comprising N fractional abundances, L is the total number of pixels, $\mathbf{w}[n]=[\,w_1[n],\ldots,w_M[n]\,]^T\sim\mathcal{N}(\mathbf{0},\sigma^2\mathbf{I}_M)$ where σ^2 is the noise variance, and $\mathbf{z}[n]$ denotes the outlier vector which only ap-

This work is supported by National Science Council (R.O.C.) under Grant NSC 99-2221-E-007-003-MY3, and by a General Research Fund of Hong Kong Research Grant Council (Project No. CUHK415509).

pears at Z pixels, i.e.,

$$\mathbf{z}[n] \neq \mathbf{0},\ n \in \mathcal{I} \triangleq \{\ell_1, ..., \ell_Z\},$$
$$\mathbf{z}[n] = \mathbf{0},\ n \in \mathcal{L} \setminus \mathcal{I},$$

where $\mathcal{L} = \{1, 2, ..., L\}$ and $\mathcal{I}$ is the set of outlier pixel indices.

Outlier-robust dimension reduction is to find an affine set representation for the corruption-free data $\mathbf{x}[n]$ from the corrupted data $\mathbf{y}[n]$ with prior knowledge of N. Some general assumptions for analysis of hyperspectral images are as follows [7]: (A1) $s_i[n] \geq 0$ for all i and n; (A2) $\sum_{i=1}^{N} s_i[n] = 1$ for all n; (A3) $\min\{L, M\} \geq N$ and $\mathbf{A}$ is of full column rank.

As has been shown in [6], the affine hull of contamination-free pixels $\mathbf{x}[n]$ is identical to that of endmembers $\mathbf{a}_1, \ldots, \mathbf{a}_N$:

$$\begin{aligned}\mathcal{A}(\mathbf{C}, \mathbf{d}) &\triangleq \text{aff}\{\mathbf{x}[1], ..., \mathbf{x}[L]\} = \text{aff}\{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \\ &= \{\mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \mid \boldsymbol{\alpha} \in \mathbb{R}^{N-1}\},\end{aligned} \tag{2}$$

for some $(\mathbf{C}, \mathbf{d}) \in \mathbb{R}^{M \times (N-1)} \times \mathbb{R}^M$, $\boldsymbol{\alpha} \in \mathbb{R}^{N-1}$. The endmember affine set parameter $(\mathbf{C}, \mathbf{d})$ has a closed-form solution for the case when $\mathbf{x}[1], \ldots, \mathbf{x}[L]$ are available. However, what we have in practice is the noisy, outlier-contaminated, observed pixel vectors $\{\mathbf{y}[n]\}_{n=1}^{L}$, and therefore obtaining an accurate estimate of $(\mathbf{C}, \mathbf{d})$ from $\{\mathbf{y}[n]\}_{n=1}^{L}$ is a challenging problem.

3. ROBUST AFFINE SET FITTING ALGORITHM

In this section, we present the RASF algorithm for estimation of $(\mathbf{C}, \mathbf{d})$ from $\{\mathbf{y}[n]\}_{n=1}^{L}$, with the prior knowledge of the number of outliers Z. Consider the RASF problem as follows:

$$\min_{\text{num}\{\mathbf{z}_1, ..., \mathbf{z}_L\} \leq Z} \left\{ \min_{\substack{\mathbf{x}_n \in \mathcal{A}(\mathbf{C}, \mathbf{d}), \forall n \\ \mathbf{C}^T\mathbf{C} = \mathbf{I}_{N-1}}} \sum_{n=1}^{L} \|\mathbf{y}[n] - \mathbf{x}_n - \mathbf{z}_n\|_2^2, \right\} \tag{3}$$

where $\text{num}\{\mathbf{z}_1, ..., \mathbf{z}_L\}$ denotes the number of nonzero vectors in $\{\mathbf{z}_1, ..., \mathbf{z}_L\}$. The objective of (3) is to seek an $(N-1)-$dimensional affine set $\mathcal{A}(\mathbf{C}, \mathbf{d})$ with the minimum projection error with respect to (w.r.t.) $\mathbf{y}[n]$ and with minimum effect of outliers $\mathbf{z}[n]$. It can be readily noted that problem (3) is nonconvex, and hence we resort to alternating optimization to handle problem (3) as follows:

(1) *Problem* (3) *w.r.t. variables* $\{\mathbf{x}_n\}_{n=1}^{L}$, $\mathbf{C}$, *and* $\mathbf{d}$*:*

$$\min_{\substack{\mathbf{x}_n \in \mathcal{A}(\mathbf{C}, \mathbf{d}),\ \mathbf{C}^T\mathbf{C} = \mathbf{I}_{N-1} \\ n=1,...,L}} \sum_{n=1}^{L} \|(\mathbf{y}[n] - \hat{\mathbf{z}}_n) - \mathbf{x}_n\|_2^2, \tag{4}$$

for any given $\{\hat{\mathbf{z}}_1, ..., \hat{\mathbf{z}}_L\}$ that satisfies $\text{num}\{\hat{\mathbf{z}}_1, ..., \hat{\mathbf{z}}_L\} \leq Z$. Following the proof in [6, Proposition 1], problem (4) can be shown to have an analytical solution given by

$$\hat{\mathbf{d}} = \frac{1}{L}\sum_{n=1}^{L}(\mathbf{y}[n] - \hat{\mathbf{z}}_n), \tag{5}$$

$$\hat{\mathbf{C}} = [\mathbf{q}_1(\mathbf{U}\mathbf{U}^T), \mathbf{q}_2(\mathbf{U}\mathbf{U}^T), ..., \mathbf{q}_{N-1}(\mathbf{U}\mathbf{U}^T)], \tag{6}$$

$$\hat{\mathbf{x}}_n = \hat{\mathbf{C}}\hat{\mathbf{C}}^T(\mathbf{y}[n] - \hat{\mathbf{z}}_n - \hat{\mathbf{d}}) + \hat{\mathbf{d}},\ n = 1, ..., L, \tag{7}$$

where $\mathbf{U} \triangleq [(\mathbf{y}[1] - \hat{\mathbf{z}}_1) - \hat{\mathbf{d}}, ..., (\mathbf{y}[L] - \hat{\mathbf{z}}_L) - \hat{\mathbf{d}}]$, and $\mathbf{q}_i(\mathbf{U}\mathbf{U}^T)$ denotes the unit-norm eigenvector associated with the ith principal eigenvalue of $\mathbf{U}\mathbf{U}^T$.

(2) *Problem* (3) *w.r.t. variables* $\{\mathbf{z}_n\}_{n=1}^{L}$*:*

$$\min_{\text{num}\{\mathbf{z}_1, ..., \mathbf{z}_L\} \leq Z} \sum_{n=1}^{L} \|(\mathbf{y}[n] - \hat{\mathbf{x}}_n) - \mathbf{z}_n\|_2^2, \tag{8}$$

for any given $\{\hat{\mathbf{x}}_n\}_{n=1}^{L} \subset \mathcal{A}(\hat{\mathbf{C}}, \hat{\mathbf{d}})$. It is trivial to see that the solution of the above problem is

$$\hat{\mathbf{z}}_n = \begin{cases} \mathbf{y}[n] - \hat{\mathbf{x}}_n, & n \in \{\hat{\ell}_1, ..., \hat{\ell}_Z\} \\ \mathbf{0}, & n \in \mathcal{L} \setminus \{\hat{\ell}_1, ..., \hat{\ell}_Z\} \end{cases} \tag{9}$$

where $\hat{\ell}_i$ is the index of the ith largest value in $(\|\mathbf{y}[1] - \hat{\mathbf{x}}_1\|, ..., \|\mathbf{y}[L] - \hat{\mathbf{x}}_L\|)$.

We generate a solution of problem (3) by handling the above two partial minimization problems alternatively until some stopping criterion is met. The pseudo-codes of the RASF algorithm for (3) are given in Table 1 (left part).

4. ESTIMATION OF THE NUMBER OF OUTLIERS USING THE RASF ALGORITHM

This section proposes the RASF-NP algorithm that equips the RASF algorithm with the capability of estimating the number of outliers by using Neyman-Pearson hypothesis testing. Let us consider the RASF problem (3) with Z being replaced by an initial guess K. Suppose that $K \geq Z$ and $\mathcal{I} \subset \{\hat{\ell}_1, ..., \hat{\ell}_K\}$. When $K = Z$, the corresponding RASF solution $\{\hat{\mathbf{x}}_n, \hat{\mathbf{z}}_n\}_{n=1}^{L}$ is exactly a local optimal approximation of $\{\mathbf{x}[n], \mathbf{z}[n]\}_{n=1}^{L}$. Hence, it can be easily inferred from (9) that when $K > Z$, the rest of $K - Z$ estimated outliers will be around zero as $\mathbf{z}[n] = 0$ for pixel indices other than those in $\mathcal{I}$. Then, by (1) the fitting error vector of the RASF problem can be expressed as

$$\mathbf{e}[n] \triangleq \mathbf{y}[n] - \hat{\mathbf{x}}_n - \hat{\mathbf{z}}_n = \boldsymbol{\mu}_n + \mathbf{w}[n],\ n = 1, ..., L, \tag{10}$$

where $\boldsymbol{\mu}_n \triangleq \mathbf{x}[n] + \mathbf{z}[n] - \hat{\mathbf{x}}_n - \hat{\mathbf{z}}_n$. There are two observations on $\mathbf{e}[n]$ given by (10) as follows:

- If $K \geq Z$, $\mathbf{e}[n]$ can be approximated to a zero-mean Gaussian random vector for all $n \neq \hat{\ell}_i,\ i = 1, \ldots, K$; i.e., $\mathbf{e}[n] \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_M), \forall n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\}$.
- If $K < Z$, there exists at least one $\mathbf{e}[n] \sim \mathcal{N}(\boldsymbol{\mu}_n, \sigma^2\mathbf{I}_M)$, where $n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\}$.

Define

$$r_n = \mathbf{e}[n]^T(\sigma^2\mathbf{I}_M)^{-1}\mathbf{e}[n],\ n = 1, ..., L, \tag{11}$$

where the noise power σ^2 is assumed to be known, and in practice, it can be estimated by multiple regression method [8]. When $K \geq Z$, it is easy to see that $r_n, \forall n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\}$ can be approximated as central Chi-square random variables $\chi^2(M)$, otherwise there exists at least one r_n for $n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\}$ being non-central Chi-square distributed $N\chi^2(M, \boldsymbol{\mu}_n)$, where M denotes the degrees of freedom. Hence, we consider the following binary hypothesis testing problem:

$$H_0\ (K \geq Z) : r_n \sim \chi^2(M),\ \forall n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\} \tag{12a}$$

$$H_1\ (K < Z) : \exists\, r_n \sim N\chi^2(M, \boldsymbol{\mu}_n),\ n \in \mathcal{L}\backslash\{\hat{\ell}_1, \ldots, \hat{\ell}_K\}. \tag{12b}$$

Since $\boldsymbol{\mu}_n$ is unknown in $N\chi^2(M, \boldsymbol{\mu}_n)$, we use Neyman-Pearson classifier rule for the above hypothesis testing problem:

$$\text{Decide } H_0 \text{ if } r_n < \eta,\ \forall n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\}, \tag{13a}$$

$$\text{Decide } H_1 \text{ if } \exists n \in \mathcal{L}\backslash\{\hat{\ell}_1, \ldots, \hat{\ell}_K\} \text{ such that } r_n > \eta, \tag{13b}$$

Table 1. The pseudo-codes of the proposed RASF and RASF-NP algorithms.

RASF Algorithm	**RASF-NP Algorithm**
Given A convergence tolerance $\varepsilon > 0$, hyperspectral data $\{\mathbf{y}[n]\}_{n=1}^{L}$, the number of endmembers N, and the number of outliers Z.	**Given** Upper and lower bounds of the number of outliers (u_p, l_o), the number of endmembers N, hyperspectral data $\{\mathbf{y}[n]\}_{n=1}^{L}$, the noise covariance matrix $\sigma^2\mathbf{I}_M$, and the false alarm rate P_{FA}.
S1. Initialize $\hat{\mathbf{z}}_1 = \cdots = \hat{\mathbf{z}}_L = \mathbf{0}$, and iteration number $k := 1$.	**S1.** Initialize $K_1 = \lceil (l_o + u_p)/2 \rceil$ (an integer), and set $i := 1$.
S2. Update the solution of inner minimization problem, $\hat{\mathbf{d}}$, $\hat{\mathbf{C}}$, and $\{\hat{\mathbf{x}}_n\}_{n=1}^{L}$ by (5), (6), and (7), respectively.	**S2.** Obtain $\{\hat{\mathbf{x}}_n, \hat{\mathbf{z}}[n]\}_{n=1}^{L}$, $(\hat{\mathbf{C}}, \hat{\mathbf{d}})$ by the RASF, and $\mathbf{e}[n]$ by (10).
S3. Update the solution of outer minimization problem $\{\hat{\mathbf{z}}_n\}_{n=1}^{L}$ by (9)	**S3.** Compute $\{r_n\}_{n=1}^{L}$ by (11) and find their maximum $\hat{r}$ by (16).
S4. Calculate fitting error $\varrho(k) = \sum_{n=1}^{L} \|\mathbf{y}[n] - \hat{\mathbf{x}}_n - \hat{\mathbf{z}}_n\|^2$.	**S4.** If $\psi(\hat{r}) > P_{FA}$, update $u_p := K_i$, otherwise update $l_o := K_i$. Then, compute $K_{i+1} = \lceil (u_p + l_o)/2 \rceil$.
S5. If $k = 1$ or $(\varrho(k-1) - \varrho(k))/\varrho(k-1) > \varepsilon$, update k by $k+1$ and go to **S2**, else output the approximate robust affine set parameters $(\hat{\mathbf{C}}, \hat{\mathbf{d}})$ and the outlier pixel indices $\hat{\mathcal{I}} = \{\hat{\ell}_1, ..., \hat{\ell}_Z\}$.	**S5.** If $K_i \neq K_{i+1}$, update $i := i + 1$ and go to **S2**, else obtain the estimate $\hat{Z} = K_i$ and the estimate $(\hat{\mathbf{C}}, \hat{\mathbf{d}})$.

where η is a parameter determined by the preassigned false alarm rate P_{FA}. Denoting the probability density function (pdf) of the central Chi-square distribution by $f_{\chi^2}(x, M)$, we define

$$\psi(r_n) \triangleq \int_{r_n}^{\infty} f_{\chi^2}(x, M)dx = 1 - \frac{\gamma(r_n/2, M/2)}{\Gamma(M/2)}, \tag{14}$$

where $\gamma(x/2, M/2)$ is the lower incomplete Gamma function [9]. Then, by Neyman-Pearson lemma [10], the optimal threshold η for problem (13) must satisfy $\psi(\eta) = P_{\rm FA}$. Although there is no closed-form expression for the inverse function of $\psi(\cdot)$, the decision rule in (13) can be equivalently formulated as

$$\text{Decide } H_0 \text{ if } \psi(r_n) < P_{\rm FA},\ \forall n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\}, \tag{15a}$$

$$\text{Decide } H_1 \text{ if } \exists n \in \mathcal{L}\backslash\{\hat{\ell}_1, \ldots,\ \hat{\ell}_K\} \text{ such that } \psi(r_n) > P_{\rm FA}. \tag{15b}$$

By the decision rule (15), we need to test at most $\psi(r_1), ..., \psi(r_L)$ to decide whether the current K is overestimation H_0 $(K \geq Z)$ or underestimation H_1 $(K < Z)$. Because $\psi(\cdot)$ is a monotone decreasing function, we can further simplify the decision rule by defining

$$\hat{r} = \max_{n \in \mathcal{L}\backslash\{\hat{\ell}_1, ..., \hat{\ell}_K\}} r_n, \tag{16}$$

Then, hypothesis testing in (15) becomes

$$\text{Decide } H_0 \text{ if } \psi(\hat{r}) > P_{\rm FA}, \tag{17a}$$

$$\text{Decide } H_1 \text{ if } \psi(\hat{r}) < P_{\rm FA}. \tag{17b}$$

Once $\psi(\hat{r})$ is evaluated, one of the above two hypotheses is decided. The pseudo codes proposed RASF-NP algorithm are also given in Table 1 (right part).

5. SIMULATION AND CONCLUSION

Monte Carlo simulations of 100 independent runs are presented to demonstrate the performance of the proposed RASF and RASF-NP methods in this section. In each run, the observed data were synthetically generated following (1) where $N = 8$ endmembers with $M = 224$ bands were selected from the U.S. geological survey (USGS) library, the number of pixels L is set to 1000, the abundance vectors were generated following Dirichlet distribution [6], and zero-mean white Gaussian noise vectors were added for different signal-to-noise ratios (SNRs), where $\text{SNR} = \sum_{n=1}^{L} \|\mathbf{x}[n]\|^2/\sigma^2 ML$. Besides, the outliers were also added to the noisy data, where the outlier indices $\ell_1, \ldots, \ell_Z$ were randomly selected from $\{1, \ldots, L\}$, and the associated outliers were generated by

$$\mathbf{z}[\ell_i] = c\boldsymbol{\kappa}_i,\ i = 1, \ldots, Z, \tag{18}$$

in which each element of $\boldsymbol{\kappa}_i$ is a zero-mean unit-variance Laplacian random variable, and c is a scalar adjusted to satisfy signal-to-outlier ratio (SOR) specification, where

$$\text{SOR} = \frac{\sum_{n=1}^{L} \|\mathbf{x}[n]\|_2^2/L}{\sum_{i=1}^{Z} \|\mathbf{z}[\ell_i]\|_2^2/Z}. \tag{19}$$

The generation of outliers using Laplacian distribution is to fulfill the belief that the outliers should be heavily tailed in distribution, which is highly peaked at zero and falls off more slowly than Gaussian distribution in the tail. When SNR$\geq$SOR, the outlier pixels $\mathbf{y}[\ell_1], ..., \mathbf{y}[\ell_Z]$ are corrupted by the outliers $\mathbf{z}[\ell_1], ..., \mathbf{z}[\ell_Z]$ more seriously than the noise; otherwise, the case of SOR$\geq$SNR means that the effects of the outliers $\mathbf{z}[\ell_1], ..., \mathbf{z}[\ell_Z]$ are smaller than the noise effects, thereby making the outlier pixels $\mathbf{y}[\ell_1], ..., \mathbf{y}[\ell_Z]$ not much different from the rest of the observed pixel vectors.

Three performance indices were used in the simulations. The distance between the true affine set $\mathcal{A}(\mathbf{C}, \mathbf{d})$ and the estimated affine set $\mathcal{A}(\hat{\mathbf{C}}, \hat{\mathbf{d}})$, denoted by $D_{\rm aff}$, for evaluation of the accuracy of the RASF is defined as

$$D_{\rm aff} = \frac{\|\mathbf{C}\mathbf{C}^T - \hat{\mathbf{C}}\hat{\mathbf{C}}^T\|_F}{\sqrt{2}} + \frac{\|\mathbf{P}_{\mathbf{C}}^{\perp}\mathbf{d} - \mathbf{P}_{\hat{\mathbf{C}}}^{\perp}\hat{\mathbf{d}}\|}{\|\mathbf{P}_{\mathbf{C}}^{\perp}\mathbf{d}\| + \|\mathbf{P}_{\hat{\mathbf{C}}}^{\perp}\hat{\mathbf{d}}\|}, \tag{20}$$

where $\|\cdot\|_F$ stands for Frobenius-norm. The first term, in range $[0, 1]$, is called the *projection F-norm* [12] and it measures the distance between the range space of $\mathbf{C}$ and $\hat{\mathbf{C}}$, and the second term in range $[0, 1]$ quantifies the error between $\mathbf{P}_{\mathbf{C}}^{\perp}\mathbf{d}$ and $\mathbf{P}_{\hat{\mathbf{C}}}^{\perp}\hat{\mathbf{d}}$. The root-mean-square (rms) spectral angle distance between the true endmembers and estimated endmembers, denoted by ϕ (in degrees), was used as an accuracy measure of EEAs [6]. The smaller the values of $D_{\rm aff}$ (or ϕ), the better the accuracy of the affine set estimates (or the endmember estimates). The computation time T (in seconds) of each algorithm (implemented in Mathworks Matlab R2008a) running in a desktop computer equipped with Core i7-930 CPU 2.80 GHz, 12GB memory is used as the computational complexity measure.

The first simulation examines the performance of the proposed RASF with different choices of the number of outliers K. Figure 1 shows the average $D_{\rm aff}$ of ASF and RASF with $K = 2\%L, 5\%L, 8\%L$ for $Z = 5\%L$, SNR=$15, 25, 35, 45, \infty$ dB and SOR=$15, 25$ dB. It can be observed that RASF algorithm perfectly identifies the true affine set when SNR=∞ and $K \geq Z$. One can also notice that RASF outperforms ASF for all the values of K under test when SOR $\leq$ SNR, and that the RASF algorithm in the case of $K \geq Z$ outperforms the case of $K < Z$ for SOR$\leq$ SNR. This implies that as long as the outlier pixels were corrupted by outliers more heavily than noise, the RASF algorithm can always mitigate the outlier effects. On the other hand, if the outliers have lower power than noise, they are simply treated as noise.

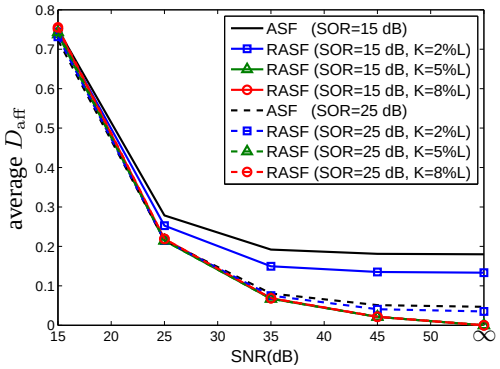

Fig. 1. Performance comparison of average D_{aff} of ASF and RASF with different preset values of K for $Z = 5\%L$, SOR$= 15, 25$ dB, and various SNR values.

The second simulation evaluates the estimation accuracy and computational efficiency of the proposed RASF-NP algorithm in comparison with the existing RSAD algorithm [5]. Table 2 shows the average number of outliers estimated by RASF-NP and RSAD algorithms with $P_{FA} = 10^{-4}, 10^{-5}, 10^{-6}$, and the average computation time T for $Z = 5\%L$, SNR=$15, 25$ dB and SOR=$10, 20$ dB. One can see that the estimated number of outliers $\hat{Z}$ obtained by the RASF-NP and RASD algorithms are quite comparable, but RASF-NP spends much less computation time than RSAD in more than one order. The reason is that RSAD repeatedly, randomly selects pixels as background and so will cost lots of computational time. Moreover, the two algorithms all perform well for SOR$<$SNR, but for SOR$>$SNR, both of them do not perform well since most outliers are treated as noise. We should emphasize again that unlike RSAD only detects outliers, the proposed RASF-NP not only detects outliers, but also provides the corruption-free affine set for dimension reduction.

The third simulation investigates the impact of the proposed RASF-NP to some existing EEAs [1, 11]. Table 3 tabulates the average ϕ of VCA, SGA, N-FINDR, AVMAX, and SVMAX, with dimension reduction using RASF-NP algorithm ($P_{FA} = 10^{-6}$) and ASF, for $Z = 5\%L$, SNR$= 15$ dB and various SOR$= 5, 7, 11, 14, 17$ dB. It can be seen that the performances of all the EEAs with ASF used improve as the SOR increases, and the RASF-NP algorithm substantially boosts the performances of all the EEAs in the presence of outliers. Besides, all the EEAs with RASF-NP used perform equally well for all SORs, implying that the RASF-NP can be in conjunction with any EEAs to provide better endmember estimates than the ASF.

In conclusion, we have presented the RASF algorithm for joint dimension reduction and outlier removal, and the RASF-NP algorithm to estimate the number of outliers, apart from robust dimension reduction and outlier detection. The proposed RASF-NP outperforms RSAD in terms of computational load by more than one order, and any EEAs in conjunction with the former will provide better endmember estimates, especially for lower SORs.

Table 2. Performance comparison of average number of the outliers $\hat{Z}$ $(\%L)$ estimated by the RASF-NP algorithm and the RSAD algorithm, and T(seconds) for true number of outliers $Z = 5\%L$, and various SNRs and SORs.

SOR (dB)	P_{FA}		SNR (dB)			
			15		25	
			RASF-NP	RSAD	RASF-NP	RSAD
10	10^{-4}	$\hat{Z}$	5.00	5.00	5.00	5.00
		T	2.56	53.37	1.39	45.61
	10^{-5}	$\hat{Z}$	5.00	5.00	5.00	5.00
		T	2.56	50.58	1.39	40.40
	10^{-6}	$\hat{Z}$	5.00	4.99	5.00	5.00
		T	2.56	53.87	1.39	38.44
20	10^{-4}	$\hat{Z}$	1.00	0.02	5.00	5.00
		T	0.93	50.44	0.75	52.91
	10^{-5}	$\hat{Z}$	1.00	0.00	5.00	4.99
		T	0.93	45.25	0.75	53.60
	10^{-6}	$\hat{Z}$	1.00	0.00	5.00	4.99
		T	0.93	40.74	0.75	54.27

Table 3. Performance comparison of average ϕ (degrees) over some existing EEAs with RASF-NP ($P_{FA} = 10^{-6}$) used and ASF used for $Z = 5\%L$, SNR$= 15$ dB, and various SORs.

EEAs	Dimension Reduction	SOR (dB)				
		5	8	11	14	17
VCA	RASF-NP	5.43	3.37	3.44	3.42	3.43
	ASF	17.31	10.46	5.66	4.18	3.74
SGA	RASF-NP	5.01	3.09	3.09	3.09	3.13
	ASF	15.39	9.58	5.11	3.55	3.22
N-FINDR	RASF-NP	5.14	3.22	3.21	3.22	3.24
	ASF	17.33	10.15	5.05	3.57	3.30
AVMAX	RASF-NP	5.18	3.21	3.20	3.23	3.25
	ASF	17.41	10.51	5.40	3.78	3.36
SVMAX	RASF-NP	5.08	3.07	3.07	3.07	3.11
	ASF	17.46	9.59	5.06	3.46	3.22

6. REFERENCES

[1] J. M. Bioucas-Dias and A. Plaza, "Hyperspectral unmixing: Geometrical, statistical, and sparse regression-based approaches," in *Proc. of SPIE - Image and Signal Processing for Remote Sensing XVI*, vol. 7830, Toulouse, France, Sept. 20, 2010, pp. 783 00A.

[2] T. Han, D. G. Goodenough, A. Dyk, and J. Love, "Detection and correction of abnormal pixels in hyperion images," in *Proc. IEEE Int. Geosci. and Remote Sens. Symp.*, vol. 3, June 25-26, 2002, pp. 1327–1330.

[3] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[4] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, 1990.

[5] Bo Du and Liangpei Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, pp. 1578–1589, May, 2011.

[6] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.

[7] N. Keshava and J. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[8] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435V2445, Aug. 2008.

[9] G. Arfken and H. Weber. *Mathematical Methods for Physicists.* Harcourt Academic Press, 2000.

[10] L. C. Ludeman. *Random Processes Filtering, Estimation, and Detection.* Wiley-Interscience Publication, 2003.

[11] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geoscience and Remote Sensing - Special Issue on Spectral Unmixing of Remotely Sensed Data*, vol. 49, no. 11, pp. 4177 - 4193, Nov. 2011.

[12] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1999.