A Comparative Study of Unsupervised Unmixing Algorithms to Detecting Anomalies in Hyperspectral Images

Li He, Jiajia Luo, Hairong Qi
Dept of Electrical Engineering and Computer Science
University of Tennessee Knoxville
Knoxville, TN 37932
Email: {lhe4,jluo9,hqi}@utk.edu

Chiman Kwan
Signal Processing Inc.
9700 Great Seneca Hwy., Rockville, MD 20850
Email: chiman.kwan@signalpro.net

Abstract—In this paper, we present a comparative study of several unsupervised unmixing algorithms to anomaly detection in hyperspectral images. The algorithms are called minimum volume constrained non-negative matrix factorization (MVC-NMF) [1], gradient descent maximum entropy (GDME) [2] and unsupervised fully constrained least squares (UFCLS) [3]. Several variants of the above algorithms were also implemented and evaluated. Actual hyperspectral image containing 4 panels and 2 small targets from the AF were used in our studies. In our experiments, MVCNMF gets the best detection results when we use the down-sampled image with full bands. Result obtained using UFCLS is close to MVCNMF and better than GDME. In addition, the speed of MVCNMF is the fastest among three methods. Updated algorithms do not provide better result than our previous work. Finally, the dimension reduction using principal component analysis (PCA) operation does affect the final results.

I. INTRODUCTION

Hyperspectral images shows vast potential for use in automatic target detection and recognition for the reason that it provides both spatial features and important information about the spectral characteristics of the materials, targets and backgrounds. [4]. The most robust class of algorithms for detection of this type of targets is arguably the one that searches the pixels of image cube for rare pixels whose information significantly differs from their surrounding pixels and local background. These targets are known as "Anomaly" in image processing and remote sensing literature.

The anomaly detection problem can be modeled using linear unmixing. Given a set of mixed hyperspectral vectors, linear mixture analysis, or linear unmixing, aims at estimating the number of reference materials, also called endmembers, their spectral signatures, and their abundance fractions. [5] It is reasonable to consider the "anomaly" is consisted by specific endmembers, and their abundance fractions should reflect the locations of the "anomaly" in the image.

An important challenge of hyperspectral unmixing is to detect endmembers and abundance without any prior knowledge. Normally, linear unmixing require complete prior knowledge of material signatures present in an image scene. Unfortunately, finding such information is generally difficult. [3] In order to solve this problem, unsupervised algorithms are developed. In this article, hence, several unsupervised unmixing algorithms are compared, include minimum volume constrained non-negative matrix factorization (MVCNMF) [1], gradient descent maximum entropy (GDME) [2], unsupervised fully constrained least squares (UFCLS) [3]. Among those algorithms, NMF was originally proposed for object recognition and has been recently applied to hyperspectral unmixing. Because NMF may suffer from a nonunique decomposition problem, a more reliable decomposition algorithm called minimum volume constrained NMF (MVCNMF) is provided [1]. In this algorithm, VCA and FCLS is hired to provide initial endmembers and abundance.
Besides, followed by Craig’s MVT idea, three related algorithms, minimum-volume enclosing simplex (MVES) [17], minimum volume simplex analysis (MVSA) [18] and simplex identification via split augmented Lagrangian (SISAL) [16], have been proposed recently. MVES and MVSA are both competitive in performance. SISAL is more efficient than other two algorithms and claimed has smaller minimum square error than others as well. [16] Therefore, SISAL is incorporated into MVCNMF and GDME as variations and tested in this article.

The organization of this paper is as follows. Section II gives general introduction of unsupervised algorithms compared in this article. Section III presents comparison results of algorithms. Finally, some conclusions are drawn in Section IV.

II. UNSUPERVISED ALGORITHMS

A. Linear Mixing Model

In solving the spectral unmixing problem, the linear mixing model (LMM) has gained significant popularity due to its effectiveness and simplicity. An LMM is valid when the endmembers are distributed as discrete patches, in which different endmembers do not interfere with each other. [6] Mathematically, the model is given by

\[ x = As + \epsilon \]  

where \( x \in \mathbb{R}^l \) is an observation vector at a single pixel with \( l \) spectral bands, \( A \in \mathbb{R}^{l \times c} \) is the material signature matrix (source matrix) whose columns, \( \{a_j\}_{j=1}^c \in \mathbb{R}^l \), correspond to the spectral signatures of different endmembers, and \( c \) is the number of endmembers. The abundance vector is denoted by \( s \in \mathbb{R}^c \), which satisfies two physical constraints, referred to as the abundance non-negative constraint, \( s_j \geq 0, j = 1, 2, \ldots, c \), and the abundance sum-to-one constraint, \( \sum_{j=1}^c s_j = 1 \). The possible errors and noises are taken into account by an \( l \)-dimensional column vector \( \epsilon \).

B. Minimum Volume Constrained Non-negative Matrix Factorization

The Minimum Volume Constrained Non-negative Matrix Factorization (MVCNMF) integrates the least squares analysis and the convex geometry model by incorporating a volume constraint into the NMF formulation. Two important facts are exploited: first, the spectral data are non-negative; second, the simplex volume determined by the endmembers is the minimum among all possible simplexes that circumscribe the data scatter space. The proposed cost function consists of two parts. One part measures the approximation error between the observed data and the reconstructions from the estimated endmembers and abundances, and the other part consists of the minimum volume constraint. We can think of these two terms serving as two forces: the external force (minimizing the approximation error) drives the estimation to move outward of the data cloud, and the internal force (minimizing the simplex volume) acts in the opposite direction by forcing the endmembers to be as close to each other as possible. [1]

Combining the goal of minimum approximation error with the volume constraint, we arrive at the following constrained optimization problem

\[
\text{minimize } f(A, s) = \frac{1}{2} \| x - As \|^2_F + \lambda J(A) \\
\text{subject to } A \geq 0, s \geq 0, 1^T_c s = 1^T_N \]

where \( 1_c(1_N) \) is a \( c(N) \)-dimensional column vector of all 1’s, and \( J(A) \) is the penalty function, calculating the simplex volume determined by the estimated endmembers. The regularization parameter \( \lambda \in \mathbb{R} \) is used to control the tradeoff between the accurate reconstruction and the volume constraint.

The volume determined by \( c \) affinely independent points \( a_1, \ldots, a_c \in \mathbb{R}^{l-1} \) is calculated by

\[ V = \frac{1}{(c-1)!} \left| \det([a_2 - a_1 | \ldots | a_c - a_1]) \right| \]

In order to overcome the problem that the determinant is not defined for a non-square matrix when these \( c \) points are in an \( l \)-space \( (c \leq l) \), Principal Component Analysis (PCA) is adopted to reduce the dimensionality of points \( a_1, \ldots, a_c \) from \( l \) to \( c - 1 \) by keeping only the \( c - 1 \) most significant principal components. Then, \( J(A) \) is formulated as

\[ J(A) = \frac{1}{2(c-1)!} \det^2([\tilde{a}_2 - \tilde{a}_1 | \ldots | \tilde{a}_c - \tilde{a}_1]) \]

where \( \tilde{a}_2 - \tilde{a}_1 | \ldots | \tilde{a}_c - \tilde{a}_1 \in \mathbb{R}^{(c-1) \times 1} \) is a low dimensional transform of \( a_2 - a_1 | \ldots | a_c - a_1 \).

For the implementation of algorithm, firstly, virtual dimensionality (VD) [19] is used to determine the number of endmembers. Secondly, VCA and fully constrained least square (FCLS) is adopted to initialize \( A \) and \( s \) matrix, which stands for endmembers and abundance. Finally, the optimization problem Eq. 2 is solved using steepest descent method.

C. Gradient Descent Maximum Entropy

The gradient descent maximum entropy (GDME) is an iterative unmixing process. Based on the endmembers extracted, a maximum entropy learning process is applied to find the abundance fractions at all image pixels. Then, the estimation errors are calculated, based on which, a new endmember is identified [2]. Hence, the endmember detection and abundance estimation are two highly correlated procedures.

GDME learning aims to solve the following minimization problem with two equality constraints corresponding to the sum-to-one constraint and the sensor measurement model. The source matrix is formed by the endmembers already determined.
minimize \( f_0(s) = \sum_{j=1}^{c} s_j \ln s_j \)
subject to \( h_0(s) = 1^T s - 1 = 0 \)
\[ h_i(s) = \sum_{j=1}^{c} a_{ij} s_j - x_i = 0, \quad i = 1, \ldots, l \]

where the domain of \( f_0 \) is the positive orthant (entropy) \( \mathbb{R}^c_+ = \{ s \in \mathbb{R}^c | s_j > 0, j = 1, \ldots, c \} \). The two equality constraints correspond to the sum-to-one constraint, \( h_0(s) \), and the sensor measurement model \( h_i(s) \). Using the method of Lagrange multipliers, Eq. 5 can easily transform from constrained minimization problem to an unconstrained problem as follow

\[
\mathcal{L}(s, \lambda, \lambda_0) = f_0(s) + \lambda^T h(s) + (\lambda_0 - 1) h_0(s)
\]

where \( \lambda \) and \( \lambda_0 \) are the Lagrange multipliers and \( h(s) \) is a column vector with the \( i \)th element being \( h_i(s) \). After getting the Lagrangian function based on Eq. 6, GDME integrated two processes to solve the convex optimization problem: (1) the endmember detection based on the global least square criterion; and (2) the maximum entropy (MaxEnt) learning algorithm to find the most possible abundance distributions for the extracted endmember set [2].

The algorithms starts with an initial endmember selected as the image pixel with the largest magnitude, then finds the most possible abundance distributions of all image pixels using the MaxEnt learning method. The pixel that generates the largest LSE between the observation and the estimation will be selected as a new endmember, which is then incorporated in the endmember set for future learning. This process is repeated until the desired number of endmembers is identified. When the SNR of a given image is below a prescribed threshold, instead of using a single pixel as the identified endmember, a set of pixels with the largest LSE are picked and then the new endmember is formed by their average value. The purpose of averaging is to reduce the effect of noise and outliers. The threshold and the number of selected pixels are determined through experiments [7].

In this paper, instead of steepest descent method used in [2], Newton’s method is used as an alternative gradient descent method because its has better performance than steepest descent.

D. Unsupervised Fully Constrained Least Squares

Unsupervised Fully Constrained Least Squares is a extension of fully constrained least squares (FCLS) approach which eliminate the a priori knowledge requirement of FCLS method. The FCLS method can be described by the following optimization problem

\[
\text{minimize } (\tilde{A} s - \tilde{x})^T (\tilde{A} s - \tilde{x}) \text{ over } s
\]
subject to \( s \geq 0 \)

In which

\[
\tilde{A} = \begin{bmatrix} \delta A \\ 1^T \end{bmatrix}, \quad \tilde{x} = \begin{bmatrix} \delta x \\ 1 \end{bmatrix}
\]

The utilization of \( \delta \) in Eq.8 controls the impact of sum-to-one constraint [3].

The implementation UFCLS algorithm is very similar to GDME algorithm. The only difference is UFCLS finds the most possible abundance distributions of all image pixels using FCLS instead of MaxEnt used in GDME.

E. Variation of unsupervised algorithms

1) Hyperspectral signal identification by minimum error:

All three unsupervised methods mentioned above didn’t implement the first step of unmixing - dimensional reduction. However, correct dimensionality reduction yielding gains in algorithm performance and complexity and in data storage. A method infers the signal subspace in hyperspectral imagery called hyperspectral signal identification by minimum error is used to update current MVCNMF and GDME algorithms. It first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense. [9]

2) Simplex identification via split augmented Lagrangian:

In MVCNMF algorithm, VCA is used to estimate initial endmembers. However, if the pure pixel assumption is not fulfilled, vertexes of simplex not necessarily stand for the endmembers. Thus, a possible attack, in the vein of the seminal ideas introduced in [15], is to fit a simplex of minimum volume to the data set. MVES and MVSA exploited this direction. However, those two algorithm is too time consuming. To tackle the hard nonconvex optimization problem, SISAL solves a sequence of nonsmooth convex subproblems, using variable splitting to obtain a constraint formulation, and then applying augmented Lagrangian technique. [16] Its faster than other two algorithms, thus we use this algorithm to replace VCA in original method as the endmember initialization step of MVCNMF and GDME.

III. COMPARISON RESULT

A. AF images for Anomaly Detection and Change Detection

The hyperspectral image we used in the test is one of five hyperspectral images provided by United States Air Force. The collection instrument is pan and tilt mounted VNIR Imaging Spectrometer with 0.45 - 0.90 um wavelengths. Four aluminum panels(Black, Green, Tan, and Silver) used in the scene, as show in Figure 1(a). In addition, there are two small targets just below four panels. The signature of four panels are shown in Figure 1(b). The original image size is 1024x800x124, in which 124 is number of hyperspectral bands. Due to the large volume of the original image, we cannot run it on our computers due to memory constraint. So two versions of original image cube are used here, one is image after PCA on hyperspectral bands (1024x800x10) and another is image after downsampling with full bands (267x342x124).
B. Comparison of unsupervised algorithms

After unmixing one can quantify a target’s abundance by summing all the mixing coefficients corresponding to its signatures. We use these values to produce a abundance map that indicates the fractional amount of the target detected at different superpixel locations. In the comparison tests, we mainly use abundance maps to compare extraction results of unsupervised unmixing algorithms.

1) AF image after PCA (1024x800x10): As shown in Fig.2(a), two panels have been identified from the image by MVCNMF. Although Fig. 2(b) and 2(c) are abundance maps with mainly background components, we can still see four panels in Fig.2(b) and one small target in Fig.2(c).

As for the GDME and UFCLS algorithm, it is extremely time consuming when the image size is large. The main reason is because we use gradient descent algorithm to calculate the abundance. For each endmember, we need to execute the gradient descent algorithm N times, where N represents the size of the image, i.e., 800 x 1024. For GDME, we set 10 as the number of endmembers. Fig.2(d), 2(e) and 2(f) shows the experimental results. We could see that GDME could only detect four panels at the same time. Although we could see the small targets and four panels from Fig.2(e), there is strong background interference. For UFCLS, also with initial 10 endmembers, two panels have been detected separately in Fig.2(g) and 2(i). We also can see one small target in Fig.2(h) with strong background interference.

2) Downsampled AF image (267x342x124): We initialized MVCNMF with up to 50 number of endmembers and let VCA algorithm to estimate the signal-to-noise ratio (SNR), we got best result, all four panels have been successfully detected in the result separately, as show in Fig. 3. Actually, this result is the best result we have among all different algorithms we tested. In Fig.3(a), part of panel has been detected without other interference, in Fig. 3(b), the panel has been detected with some of background at the bottom of image. In 3(c) and 3(d), rest two panels have been detected successfully without noise. Fig. 3(e) shows the signature of four detected endmembers. One of signatures higher than others is because it includes some background information. One of the expected advantages of the MVCNMF would be the resistance to the presence of noise, [1] We can see from detected abundance maps that comparing to other results, the resistance to noise quality of MVCNMF is the best.

For GDME, three panels and one small target could be detected with some background interference, as shown in Fig.4(a), 4(b), 4(c) and 4(d). We have tested up 10 to 45 for each algorithm. However, no matter which number we choose, we get similar experimental results.

The experiment result of UFCLS method is between the MVCNMF and GDME. Two panels are successfully obtained in Fig. 5(b) and 5(c) with little noise. Another panel also detected in Fig.5(a), although its not very obvious. This result is more clear than result of GDME, although its not as clear as result of MVCNMF. However, we need to set the initial number of endmembers to 80 to get this result. The time taken
of UFCLS is several times more than MVCNMF, while the result is not as satisfied as MVCNMF.

From the test result of upon three methods, there is obvious difference with the test result on image after PCA. Hence, we might conclude that PCA does influence the experimental results.

For the HySime updated MVCNMF method, Fig.6(f) shows the mean square error, projection error and noise power of Hysime method. Hysime identified 47 endmembers as the initial endmembers to initialize MVCNMF, this number is further reduced by VCA to 20 (set manually and 20 is experienced value). Finally we obtain 20 estimated endmembers and abundance maps, among which Fig.6(a), 6(b) and 6(c) shows the best three of them. Three panels have been successfully detected with coexist of other panels and relatively large background noise, comparing to the original MVCNMF method.

For the HySime updated GDME method, by observing the Fig.6(d) and 6(e), we could find out that the experimental results are better than our previous work. We could detect the leftmost targets as well as the two targets to the right. In addition, we could find out that the second target from the left is the most difficult one to detect, since it is highly mixed.

In our experiment, we use SISAL to replace VCA in MVCNMF and to extract the endmember in GDME. Experimental results are shown in Fig.7. As observed, this method does not provide better result than our previous work. In other words, the signatures are still not accurate enough for us to get the good unmixing result.

IV. CONCLUSION

In our experiment, surprisingly, MVCNMF gets the best anomaly detection results between all unsupervised algorithms when we use the down-sampled image with full bands. In addition, MVCNMF runs fastest among three algorithms. Although UFCLS’s result is close to MVCNMF, the time it takes is far too long than MVCNMF method. As for the GDME, the test result is not as good as MVCNMF and UFCLS. In addition, The PCA operation does affect the final results when we use the MVCNMF method.

We tested the HySime and SISAL improved MVCNMF and GDME algorithm. However, the performance did not exceed the original algorithm, except HySime updated GDME is better. In the future, the MVCNMF and GDME method can be improved by more adaptive optimization algorithm instead of gradient descend.

ACKNOWLEDGMENT

This work was supported in part by Air Force Office of Scientific Research under contract FA9550-09-C-0162.
REFERENCES


